Near-Optimal Φ-Regret Learning in Extensive-Form Games

Ioannis Anagnostides[1], Gabriele Farina[2], and Tuomas Sandholm[3]

[1,3]Carnegie Mellon University
[2]Meta AI
[3]Strategy Robot, Inc.
[3]Optimized Markets, Inc.
[3]Strategic Machine, Inc.
{ianagnos,gfarina}@cs.cmu.edu, and sandholm@cs.cmu.edu

Abstract

In this paper, we establish efficient and uncoupled learning dynamics so that, when employed by all players in multiplayer perfect-recall imperfect-information extensive-form games, the *trigger regret* of each player grows as $O(\log T)$ after T repetitions of play. This improves exponentially over the prior best known trigger-regret bound of $O(T^{1/4})$, and settles a recent open question by Bai et al. (2022). As an immediate consequence, we guarantee convergence to the set of *extensive-form correlated equilibria* and *coarse correlated equilibria* at a near-optimal rate of $\frac{\log T}{T}$.

Building on prior work, at the heart of our construction lies a more general result regarding fixed points deriving from rational functions with *polynomial degree*, a property that we establish for the fixed points of *(coarse) trigger deviation functions*. Moreover, our construction leverages a refined *regret circuit* for the convex hull, which—unlike prior guarantees—preserves the *RVU property* introduced by Syrgkanis et al. (NIPS, 2015); this observation has an independent interest in establishing near-optimal regret under learning dynamics based on a CFR-type decomposition of the regret.

1 Introduction

A primary objective of artificial intelligence is the design of agents that can adapt effectively in complex and nonstationary multiagent environments—modeled as *general-sum games*. Multiagent decision making often occurs in a decentralized fashion, with each agent only obtaining information about its own reward function, and the goal is to *learn* how to play the game through repeated interactions. This begs the question: *How do we measure the performance of a learning agent?* A popular metric commonly used is that of *external regret* (or simply regret). However, external regret can be a rather weak benchmark: a no-external-regret agent could still incur substantial regret under simple in-hindsight "transformations" of its behavior—*e.g.*, consistently switching from an action a to a different action a' [Gordon et al., 2008].

A more general metric is Φ *regret* [Hazan and Kale, 2007, Rakhlin et al., 2011, Stoltz and Lugosi, 2007, Greenwald and Jafari, 2003], parameterized by a set deviations Φ. From a game-theoretic standpoint, the importance of this framework is that different choices of Φ lead to different types of equilibria [Greenwald and Jafari, 2003, Stoltz and Lugosi, 2007]. For example, one such celebrated result guarantees that *no-internal-regret* players converge—in terms of empirical frequency of play—to the set of *correlated equilibria (CE)* [Foster and Vohra, 1997, Hart and Mas-Colell, 2000]. This brings us to the following central question:

What are the best performance guarantees when no-Φ-regret learners are playing in multiplayer general-sum games?

Special cases of this question have recently received considerable attention in the literature [Daskalakis et al., 2011, Rakhlin and Sridharan, 2013a,b, Syrgkanis et al., 2015, Foster et al., 2016, Wei and Luo, 2018, Chen and Peng, 2020, Hsieh et al., 2021, Daskalakis et al., 2021, Daskalakis and Golowich, 2022, Anagnostides et al., 2022a, Piliouras et al., 2021]. In particular, Daskalakis et al. [2021] were the first to establish $O(\text{polylog}(T))$ external regret bounds for normal-form games,[1] and subsequent work extended those results to internal regret [Anagnostides et al., 2022a]; those guarantees, applicable when *all* players employ specific learning dynamics, improve exponentially over what is possible when a player is facing a sequence of adversarially produced utilities—the canonical consideration in online learning. However, much less is known about Φ-regret learning beyond normal-form games.

One important application revolves around learning dynamics for *extensive-form correlated equilibria (EFCE)* [Von Stengel and Forges, 2008, Gordon et al., 2008, Celli et al., 2020, Morrill et al., 2021a, Anagnostides et al., 2022b, Morrill et al., 2021b, Bai et al., 2022a, Song et al., 2022]. Indeed, a particular instantiation of Φ regret, referred to as *trigger regret*, is known to drive the rate of convergence to EFCE. Incidentally, minimizing trigger regret lies at the boundary of Φ-regret minimization problems that are known to be computationally tractable in extensive-form games. In this context, prior work established $O(T^{1/4})$ per-player trigger regret bounds [Anagnostides et al., 2022b], thereby leaving open the possibility of obtaining near-optimal rates for EFCE; that question was also recently posed by Bai et al. [2022a].

[1] With a slight abuse of notation, we use the $O(\cdot)$ notation in our introduction to suppress parameters that depend (polynomially) on the description of the game.

1.1 Our Contributions

Our main contribution is to establish the first uncoupled learning dynamics with near-optimal per-player trigger regret guarantees:

Theorem 1.1 (Informal; precise version in Theorem 3.10)**.** *There exist uncoupled and efficient learning dynamics so that the trigger regret of each player grows as $O(\log T)$ after T repetitions of play.*

This improves exponentially over the $O(T^{1/4})$ bounds obtained in prior work [Celli et al., 2020, Farina et al., 2021, Anagnostides et al., 2022b], and settles an open question recently posed by Bai et al. [2022b]. As an immediate consequence, given that trigger regret drives the rate of convergence to EFCE (Theorem 2.3), we obtain the first near-optimal rates to EFCE.

Corollary 1.2. *There exist uncoupled and efficient learning dynamics converging to EFCE at a near-optimal rate of $\frac{\log T}{T}$.*

Overview of our techniques Our construction leverages the template of Gordon et al. [2008] for minimizing Φ regret (Algorithm 1). In particular, we follow the regret decomposition approach of Farina et al. [2021] to construct an external regret minimizer *for the set of deviations* corresponding to *trigger deviation functions.* A key difference is that we instantiate each regret minimizer using the recent algorithm of Farina et al. [2022], namely `LRL-OFTRL`, which is based on *optimistic follow the regularizer leader (OFTRL)* [Syrgkanis et al., 2015] under logarithmic regularization; `LRL-OFTRL` guarantees suitable *RVU bounds* [Syrgkanis et al., 2015] for each "local" regret minimizer.

To combine those local regret minimizers into a global one for the set of trigger deviations that still enjoys a suitable RVU bound, we provide a refined guarantee for the "regret circuit" of the convex hull (Proposition 3.3), which ensures that the RVU property is preserved along the construction. Incidentally, this simple observation can be used to obtain the first near-optimal regret guarantees for algorithms based on a CFR-type decomposition of the regret [Zinkevich et al., 2007].

The next key step relates to the behavior of the fixed points of trigger deviation functions. (Fixed points are at heart of all known constructions for minimizing Φ regret.) More precisely, to convert the RVU property from the space of deviations to the actual space of the player's strategies, we show that it suffices that the fixed points deriving from trigger deviation functions can be expressed as a rational function with a polynomial degree (Lemma 3.5). Importantly, we prove this property for the fixed points of trigger deviation functions (Proposition 3.7), thereby leading to Theorem 1.1; the last part of our analysis builds on a technique developed for obtaining $O(\log T)$ swap regret in normal-form games [Anagnostides et al., 2022c]. We also obtain slightly improved guarantees for extensive-form *coarse* correlated equilibria (EFCCE) [Farina et al., 2020], a relaxation of EFCE that is attractive due to its reduced per-iteration complexity compared to EFCE.

Finally, we verify our theory through experiments on several benchmark extensive-form games in Section 4.

1.2 Further Related Work

Φ regret has received extensive attention as a solution concept in the literature since it strengthens and unifies many common measures of performance in online learning (*e.g.*, see [Hazan and Kale, 2007, Rakhlin et al., 2011, Stoltz and Lugosi, 2007, Greenwald and Jafari, 2003, Marks, 2008, Piliouras et al., 2022]). This framework has been particularly influential in game theory given that no-Φ-regret learning outcomes are known to converge to different equilibrium concepts, depending on the richness of the set of deviations Φ. For example, when Φ includes *all constant transformations*—reducing to external regret—no-regret learning outcomes are known to converge to *coarse correlated equilibria (CCE)* [Moulin and Vial, 1978], a relaxation of CE [Aumann, 1974]. Unfortunately, CCE is understood to be a weak equilibrium concept, potentially prescribing irrational behavior [Dekel and Fudenberg, 1990, Viossat and Zapechelnyuk, 2013, Giannou et al., 2021]. This motivates enlarging the set of deviaitons Φ, thereby leading to stronger—and arguably more plausible—equilibrium concepts. Indeed, the framework of Φ regret has been central in the development of the first uncoupled no-regret learning dynamics for EFCE [Celli et al., 2020, Farina et al., 2021] (see also [Morrill et al., 2021a,b, Zhang, 2022]).[2]

Our paper lies at the interface of the aforedescribed literature with a recent line of work that strives for improved regret guarantees when specific learning dynamics are in place; this allows bypassing the notorious $\Omega(\sqrt{T})$ lower bounds applicable under an adversarial sequence of utilities [Cesa-Bianchi and Lugosi, 2006]. The later line of work was pioneered by Daskalakis et al. [2011], and has been thereafter extended along several lines [Rakhlin and Sridharan, 2013a,b, Syrgkanis et al., 2015, Chen and Peng, 2020, Daskalakis et al., 2021, Daskalakis and Golowich, 2022, Piliouras et al., 2021], incorporating partial or noisy information feedback [Foster et al., 2016, Wei and Luo, 2018, Hsieh et al., 2022], and more recently, general Markov games [Erez et al., 2022, Zhang et al., 2022].

A key reference point for our paper is the work of Anagnostides et al. [2022b], which established $O(T^{1/4})$ trigger regret bounds through *optimistic hedge*. Specifically, building on [Chen and Peng, 2020], they showed *multiplicative stability* of the fixed points associated with EFCE. While those works operate in the full information model, recent papers have also developed dynamics converging to EFCE under bandit feedback [Bai et al., 2022a, Song et al., 2022].

2 Preliminaries

In this section, we introduce our notation and basic background on online learning and extensive-form games. For a more comprehensive treatment on those subjects, we refer to [Cesa-Bianchi and Lugosi, 2006] and [Leyton-Brown and Shoham, 2008], respectively.

Notation We denote by $\mathbb{N} = \{1, 2, \dots\}$ the set of natural numbers. We use the variable i with a subscript to index a player, and t with a superscript to indicate the (discrete) time. To access the r-th coordinate of a d-dimensional vector $\boldsymbol{x} \in \mathbb{R}^d$, for some index $r \in [\![d]\!] := \{1, 2, \dots, d\}$, we use the symbol $\boldsymbol{x}[r]$.

2.1 Online Learning and Regret

Let $\mathcal{X} \subseteq [0, 1]^d$ be a nonempty convex and compact set, for $d \in \mathbb{N}$. In the framework of online learning, a learner (or a player), denoted by $\mathfrak{R}$, interacts with the environment at time $t \in \mathbb{N}$ via the following subroutines.

[2] While there are other methods for efficiently computing EFCE [Dudík and Gordon, 2009, Huang and von Stengel, 2008], approaches based on uncoupled no-regret learning typically scale significantly better in large games.

- $\mathfrak{R}.\textsc{NextStrategy}()$: The learner outputs its next strategy $\boldsymbol{x}^{(t)} \in \mathcal{X}$ based on its internal state; and

- $\mathfrak{R}.\textsc{ObserveUtility}(\boldsymbol{u}^{(t)})$: The learner receives a feedback from the environment in the form of a utility vector $\boldsymbol{u}^{(t)} \in \mathbb{R}^d$.

The canonical measure of performance in online learning is the notion of *regret*, denoted by Reg^T, defined for a fixed time horizon $T \in \mathbb{N}$ as

$$\max_{\boldsymbol{x}^\star \in \mathcal{X}} \left\{ \sum_{t=1}^{T} \langle \boldsymbol{x}^\star, \boldsymbol{u}^{(t)} \rangle \right\} - \sum_{t=1}^{T} \langle \boldsymbol{x}^{(t)}, \boldsymbol{u}^{(t)} \rangle. \tag{1}$$

In words, the performance of the learner is compared to the performance of playing an optimal *fixed strategy* in hindsight. We will say that the agent has *no-regret* if $\mathrm{Reg}^T = o(T)$, under *any* sequence of observed utilities.

A much more general performance metric is Φ *regret*, parameterized by a set of *transformations* $\Phi : \mathcal{X} \to \mathcal{X}$. Namely, Φ-regret Reg_Φ^T—for a time horizon $T \in \mathbb{N}$—is defined as

$$\sup_{\phi^\star \in \Phi} \left\{ \sum_{t=1}^{T} \langle \phi^\star(\boldsymbol{x}^{(t)}), \boldsymbol{u}^{(t)} \rangle \right\} - \sum_{t=1}^{T} \langle \boldsymbol{x}^{(t)}, \boldsymbol{u}^{(t)} \rangle. \tag{2}$$

External regret (1) is simply a special case of (2) when Φ includes all possible *constant transformations*, but Φ regret can be much more expressive. A celebrated game-theoretic motivation for Φ regret stems from the fact that when all players employ suitable Φ-regret minimizers, the dynamics converge to different notions of *correlated equilibria*, well-beyond *coarse correlated equilibria* [Foster and Vohra, 1997, Stoltz and Lugosi, 2007, Hart and Mas-Colell, 2000, Celli et al., 2020].

From external to Φ regret As it turns out, there is a general template for minimizing Φ regret due to Gordon et al. [2008]. In particular, their algorithm assumes access to the following.

1. A *no-external-regret* minimizer $\mathfrak{R}_\Phi$ operating *over the set of transformations* Φ; and

2. A *fixed point oracle* $\textsc{FixedPoint}(\phi)$ that, for any $\phi \in \Phi$, computes a fixed point $\boldsymbol{x} \in \mathcal{X}$, under the assumption that such a point indeed exists.

Based on those ingredients, Gordon et al. [2008] were able to construct a regret minimization algorithm $\mathfrak{R}$ with sublinear Φ regret, as illustrated in Algorithm 1. Specifically, $\mathfrak{R}$ determines its next strategy by first obtaining the strategy $\phi^{(t)}$ of $\mathfrak{R}_\Phi$ (Line 2), and then outputing any fixed point of $\phi^{(t)}$ (Line 3). Then, upon receiving the utility vector $\boldsymbol{u}^{(t)} \in \mathbb{R}^d$, $\mathfrak{R}$ forwards as input to $\mathfrak{R}_\Phi$ the utility function $\phi \mapsto \langle \boldsymbol{u}^{(t)}, \phi(\boldsymbol{x}^{(t)}) \rangle$. We will assume that Φ contains *linear transformations*, in which case that utility can be represented as $\boldsymbol{U}^{(t)} := \boldsymbol{u}^{(t)} \otimes \boldsymbol{x}^{(t)}$ (Line 6), where $\otimes$ denotes the outer product of the two vectors. This algorithm enjoys the following guarantee.

Theorem 2.1 ([Gordon et al., 2008])**.** *Let* Reg^T *be the external regret of* $\mathfrak{R}_\Phi$*, and* Reg_Φ^T *be the* Φ*-regret of* $\mathfrak{R}$*. Then, for any* $T \in \mathbb{N}$*,*

$$\mathrm{Reg}^T = \mathrm{Reg}_\Phi^T.$$

Algorithm 1: Φ-Regret Minimizer [Gordon et al., 2008]

```
Data: An external regret minimizer $\mathfrak{R}_\Phi$ for $\Phi$
function NextStrategy()
    $\phi^{(t)} \leftarrow \mathfrak{R}_\Phi$.NextStrategy()
    $\boldsymbol{x}^{(t)} \leftarrow$ FixedPoint($\phi^{(t)}$)
    return $\boldsymbol{x}^{(t)}$
─────────────────────────────
function ObserveUtility($\boldsymbol{u}^{(t)}$)
    Construct the utility $\boldsymbol{U}^{(t)} \leftarrow \boldsymbol{u}^{(t)} \otimes \boldsymbol{x}^{(t)}$
    $\mathfrak{R}_\Phi$.ObserveUtility($\boldsymbol{U}^{(t)}$)
```

No-Regret learning in games The main focus of our paper is about the behavior of no-regret learning dynamics when employed by all players in n-player games. More precisely, the strategy set of each player $i \in [\![n]\!]$ is a nonempty convex and compact set $\mathcal{X}_i$. Further, the utility function $u_i : \bigtimes_{i'=1}^{n} \mathcal{X}_{i'} \to \mathbb{R}$ of each player $i \in [\![n]\!]$ is multilinear, so that for any $\boldsymbol{x}_{-i} \coloneqq (\boldsymbol{x}_1, \dots, \boldsymbol{x}_{i-1}, \boldsymbol{x}_{i+1}, \boldsymbol{x}_n)$, $u_i(\boldsymbol{x}) \coloneqq \langle \boldsymbol{x}_i, \boldsymbol{u}_i(\boldsymbol{x}_{-i}) \rangle$.

In this context, learning procedures work as follows. At every iteration $t \in \mathbb{N}$ each player $i \in [\![n]\!]$ commits to a strategy $\boldsymbol{x}_i^{(t)} \in \mathcal{X}_i$, and subsequently receives as feedback the utility corresponding to the other players' strategies at time t: $\boldsymbol{u}_i^{(t)} \coloneqq \boldsymbol{u}_i(\boldsymbol{x}_{-i}^{(t)})$. It is assumed that players use no-regret learning algorithms to adapt to the behavior of the other players, leading to *uncoupled* learning dynamics, in the sense that players do not use information about other players' utilities [Hart and Mas-Colell, 2000, Daskalakis et al., 2011]. For convenience, and without any loss, we assume that $\|\boldsymbol{u}_i^{(t)}\|_\infty \leq 1$, for $i \in [\![n]\!]$ and $t \in \mathbb{N}$.

2.2 Background on EFGs

An *extensive-form game (EFG)* is played on a rooted and directed tree with node-set $\mathcal{H}$. Every *decision (non-terminal) node* $h \in \mathcal{H}$ is uniquely associated with a player who selects an action from a finite and nonempty set $\mathcal{A}_h$. By convention, the set of players includes a fictitious "chance" player c that acts according to a fixed distribution. The set of *leaves (terminal) nodes* $\mathcal{Z} \subseteq \mathcal{H}$ corresponds to different outcomes of the game. Once the game reaches a terminal node $z \in \mathcal{Z}$, every player $i \in [\![n]\!]$ receives a payoff according to a (normalized) utility function $u_i : \mathcal{Z} \to [-1, 1]$.

In an imperfect-information EFG, the decision nodes of each player $i \in [\![n]\!]$ are partitioned into *information sets* $\mathcal{J}_i$, inducing a partially ordered set $(\mathcal{J}_i, \prec)$. For an information set $j \in \mathcal{J}_i$ and an action $a \in \mathcal{A}_j$, we let $\sigma \coloneqq (j, a)$ be the *sequence* of i's actions encountered from the root of the tree until (and including) action a; we use the special symbol $\varnothing$ to denote the *empty sequence*. The set of i's sequences is denoted by $\Sigma_i \coloneqq \{(j, a) : j \in \mathcal{J}_i, a \in \mathcal{A}_j\} \cup \{\emptyset\}$. We also let $\Sigma_i^* \coloneqq \Sigma_i \setminus \{\emptyset\}$ and $\Sigma_j \coloneqq \{\sigma \in \Sigma_i : \sigma \succeq j\}$. We will write σ_j to represent the *parent sequence* of an information set $j \in \mathcal{J}_i$; namely, the last sequence before reaching j, or $\varnothing$ if j is a *root information set.*

Sequence-form strategies The strategy of a player specifies a probability distribution for every information set encountered in the tree. Assuming *perfect-recall*—players never forget acquired information—a strategy can be represented via the *sequence-form strategy polytope* $\mathcal{Q}_i \subseteq \mathbb{R}_{\geq 0}^{|\Sigma_i|}$, defined as

$$\mathcal{Q}_i := \left\{ \boldsymbol{q}_i \in \mathbb{R}_{\geq 0}^{|\Sigma_i|} : \boldsymbol{q}_i[\varnothing] = 1, \boldsymbol{q}_i[\sigma_j] = \sum_{a \in \mathcal{A}_j} \boldsymbol{q}_i[(j,a)], \forall j \in \mathcal{J}_i \right\}.$$

Further, we let $\Pi_i := \mathcal{Q}_i \cap \{0,1\}^{|\Sigma_i|}$ be the set of *deterministic* sequence-form strategies. Analogously, one can define the sequence-form polytope $\mathcal{Q}_j$ rooted at information set $j \in \mathcal{J}_i$. We also use $\|\mathcal{Q}_i\|_1$ to denote the maximum ℓ_1-norm of a vector $\boldsymbol{q}_i \in \mathcal{Q}_i$. Finally, we denote by $\mathfrak{D}_i$ the depth of i's subtree.

Trigger deviations and EFCE To formalize the connection between EFCE and the framework of Φ-regret, we introduce *trigger deviation functions.*

Definition 2.2 ([Farina et al., 2021]). A trigger deviation function with respect to a trigger sequence $\hat{\sigma} = (j,a) \in \Sigma_i^*$ and a continuation strategy $\hat{\boldsymbol{\pi}}_i \in \Pi_j$ is any linear mapping $f : \mathbb{R}^{|\Sigma_i|} \to \mathbb{R}^{|\Sigma_i|}$ such that

- $f(\boldsymbol{\pi}_i) = \boldsymbol{\pi}_i$ for all $\boldsymbol{\pi}_i \in \Pi_i$ such that $\boldsymbol{\pi}_i[\hat{\sigma}] = 0$;
- Otherwise, for all $\boldsymbol{\pi}_i \in \Pi_i$,

$$f(\boldsymbol{\pi}_i)[\sigma] = \begin{cases} \boldsymbol{\pi}_i[\sigma] & \text{if } \sigma \not\succeq j, \\ \hat{\boldsymbol{\pi}}_i[\sigma] & \text{if } \sigma \succeq j. \end{cases}$$

We denote by Ψ_i the convex hull of all trigger deviation functions—over all trigger sequences and deterministic continuation strategies; Ψ_i-regret is referred to as *trigger regret.* In an *extensive-form correlated equilibrium (EFCE)* [Von Stengel and Forges, 2008] no trigger deviation by any player can improve the utility of that player, leading to the following connection.

Theorem 2.3 ([Farina et al., 2021]). *If each player i incurs trigger regret $\mathrm{Reg}^T_{\Psi_i}$ after T repetitions of the game, the average product distribution of play is a $\frac{1}{T}\max_{i \in [\![n]\!]} \mathrm{Reg}^T_{\Psi_i}$-approximate EFCE.*

Moreover, extensive-form *coarse* correlated equilibria (EFCCE) [Farina et al., 2020] are defined analogously based on *coarse trigger deviations* $\tilde{\Psi}_i$; the difference is that in EFCCE the player decides whether to follow the recommendation *before* actually seeing the recommendation at that information set (see Appendix A for the definition and specific examples).

3 Near-Optimal Learning for EFCE

In this section, we establish our main result: efficient learning dynamics with $O(\log T)$ per-player trigger regret; this is made precise in Theorem 3.10, the informal version of which was stated earlier in Theorem 1.1. All the proofs from this section are deferred to Appendix B.

3.1 Regret Minimizer for Trigger Deviations

Our construction for minimizing trigger regret uses the general template of Gordon et al. [2008] (Algorithm 1), and in particular, the approach of Farina et al. [2021] in order to construct an external regret minimizer for the set Ψ_i (a similar approach also applies for the set of coarse trigger deviations $\tilde{\Psi}_i$). More precisely, that construction leverages one separate regret minimizer $\mathfrak{R}_{\hat{\sigma}}$ for every possible *trigger sequence* (recall Definition 2.2) $\hat{\sigma} \in \Sigma_i^*$. In Particular, $\mathfrak{R}_{\hat{\sigma}}$, with $\hat{\sigma} = (j, a)$, is—after performing an affine transformation—operating over sequence-form vectors $\boldsymbol{q}_{\hat{\sigma}} \in \mathcal{Q}_j$ (rooted at information set $j \in \mathcal{J}_i$). Then, those regret minimizers are combined using a regret minimizer $\mathfrak{R}_\triangle$ operating over the simplex $\Delta(\Sigma_i^*)$. The first key ingredient in our construction is the use of a *logarithmic regularizer*.

In particular, we instantiate each regret minimizer with `LRL-OFTRL`, a recent algorithm due to Farina et al. [2022]. `LRL-OFTRL` is an instance of *optimistic follow the regularizer leader (OFTRL)* [Syrgkanis et al., 2015] with logarithmic regularization; the main twist is that `LRL-OFTRL` operates over an appropriately *lifted space*. For our purposes, we first modify [Farina et al., 2022, Proposition 2 and Corollary 1] to obtain a suitable RVU bound for each regret minimizer $\mathfrak{R}_{\hat{\sigma}}$ instantiated with `LRL-OFTRL`, for each $\hat{\sigma} \in \Sigma_i^*$.

Proposition 3.1. *Fix any $\hat{\sigma} \in \Sigma_i^*$, and let $\operatorname{Reg}_{\hat{\sigma}}^T$ be the regret of $\mathfrak{R}_{\hat{\sigma}}$ up to time $T \geq 2$. For any $\eta \leq \frac{1}{256\|\mathcal{Q}_i\|_1}$, $\max\{0, \operatorname{Reg}_{\hat{\sigma}}^T\}$ can be upper bounded by*

$$\frac{2|\Sigma_i| \log T}{\eta} + 16\eta\|\mathcal{Q}_i\|_1^2 \sum_{t=1}^{T-1} \|\boldsymbol{U}_i^{(t+1)} - \boldsymbol{U}_i^{(t)}\|_\infty^2 - \frac{1}{512\eta} \sum_{t=1}^{T-1} \|\boldsymbol{q}_{\hat{\sigma}}^{(t+1)} - \boldsymbol{q}_{\hat{\sigma}}^{(t)}\|_{\boldsymbol{q}_{\hat{\sigma}}^{(t)},\infty}^2. \tag{3}$$

A few remarks are in order. First, we recall that $\boldsymbol{U}_i^{(t)} := \boldsymbol{u}_i^{(t)} \otimes \boldsymbol{x}_i^{(t)}$, in accordance to Line 6 of Algorithm 1. Also, $\eta > 0$ denotes the (time-invariant) *learning rate* of `LRL-OFTRL`. Furthermore, for $\hat{\sigma} = (j, a) \in \Sigma_i^*$, in Proposition 3.1 we used the notation

$$\|\boldsymbol{q}_{\hat{\sigma}}^{(t+1)} - \boldsymbol{q}_{\hat{\sigma}}^{(t)}\|_{\boldsymbol{q}_{\hat{\sigma}}^{(t)},\infty}^2 := \max_{\sigma \in \Sigma_j} \left| 1 - \frac{\boldsymbol{q}_{\hat{\sigma}}^{(t+1)}[\sigma]}{\boldsymbol{q}_{\hat{\sigma}}^{(t)}[\sigma]} \right|.$$

Proposition 3.1 establishes an RVU bound [Syrgkanis et al., 2015], but with two important refinements. First, the bound applies to $\max\{0, \operatorname{Reg}_{\hat{\sigma}}^T\}$, instead of $\operatorname{Reg}_{\hat{\sigma}}^T$, ensuring that (3) is nonnegative. Further, the local norm appearing in (3) will also be crucial for our argument in the sequel (Lemma 3.5). Next, similarly to Proposition 3.1, we obtain the following regret bound for $\mathfrak{R}_\triangle$, the regret minimizer "mixing" over all $\{\mathfrak{R}_{\hat{\sigma}}\}_{\hat{\sigma} \in \Sigma_i^*}$.

Proposition 3.2. *Let $\operatorname{Reg}_\triangle^T$ be the regret of $\mathfrak{R}_\triangle$ up to time $T \geq 2$. For any $\eta_\triangle \leq \frac{1}{512|\Sigma_i|}$, $\max\{0, \operatorname{Reg}_\triangle^T\}$ can be upper bounded by*

$$\frac{2|\Sigma_i| \log T}{\eta_\triangle} + 16\eta_\triangle|\Sigma_i|^2 \sum_{t=1}^{T-1} \|\boldsymbol{u}_\triangle^{(t+1)} - \boldsymbol{u}_\triangle^{(t)}\|_\infty^2 - \frac{1}{512\eta_\triangle} \sum_{t=1}^{T-1} \|\boldsymbol{\lambda}_i^{(t+1)} - \boldsymbol{\lambda}_i^{(t)}\|_{\boldsymbol{\lambda}_i^{(t)},\infty}^2.$$

Here, we used the notation $\boldsymbol{u}_\triangle^{(t)}[\hat{\sigma}] := \langle \boldsymbol{X}_{\hat{\sigma}}^{(t)}, \boldsymbol{U}_i^{(t)} \rangle$, where $\boldsymbol{X}_{\hat{\sigma}}^{(t)}$ is the output of $\mathfrak{R}_{\hat{\sigma}}$, for each $\hat{\sigma} \in \Sigma_i^*$; that is, $\boldsymbol{X}_{\hat{\sigma}}^{(t)} \in \mathbb{R}^{|\Sigma_i| \times |\Sigma_i|}$ transforms sequence-form vectors based on the continuation strategy $\boldsymbol{q}_{\hat{\sigma}}^{(t)}$ below the trigger sequence $\hat{\sigma}$ (recall Definition 2.2). We are now ready to use Theorem 2.1 to obtain a bound for Ψ_i-regret.

Proposition 3.3. *For any $T \in \mathbb{N}$,*

$$\max\{0, \operatorname{Reg}_{\Psi_i}^T\} \leq \max\{0, \operatorname{Reg}_\triangle^T\} + \sum_{\hat{\sigma} \in \Sigma_i^*} \max\{0, \operatorname{Reg}_{\hat{\sigma}}^T\}.$$

This uses the *regret circuit* for the convex hull [Farina et al., 2019a] to combine all the regret minimizers $\{\mathfrak{R}_{\hat{\sigma}}\}_{\hat{\sigma} \in \Sigma_i^*}$ via $\mathfrak{R}_{\triangle}$ into an external regret minimizer for the set Ψ_i; by virtue of Theorem 2.1, the external regret of the induced algorithm is equal to the Ψ_i-regret ($\mathrm{Reg}_{\Psi_i}^T$) of player i. There is, however, one crucial twist: the guarantee of Farina et al. [2019a] would give a bound in terms of $\max_{\hat{\sigma} \in \Sigma_i^*} \mathrm{Reg}_{\hat{\sigma}}^T$, instead of $\sum_{\hat{\sigma} \in \Sigma_i^*} \mathrm{Reg}_{\hat{\sigma}}^T$; this is problematic for obtaining near-optimal rates as *it breaks the RVU property over the convex hull.* In general, it is not clear how to bound the maximum of the regrets by their sum since (external) regret *can be negative.* This is, in fact, a recurrent obstacle encountered in this line of work [Syrgkanis et al., 2015], and it is precisely the reason why approaches based on regret decomposition—in the spirit of CFR [Zinkevich et al., 2007]—failed to bring rates better than $T^{-3/4}$ [Farina et al., 2019b]. Proposition 3.3 circumvents those obstacles by establishing bounds in terms of *nonnegative measures of regret.*

Remark 3.4 (Near-optimal regret via CFR-type algorithms). An important byproduct of our techniques, and in particular Proposition 3.3 along with RVU bounds for nonnegative measures of regret [Anagnostides et al., 2022c], is the first near-optimal $O(\log T)$ regret bound for CFR-type algorithms in general games, a question that has been open even in zero-sum games.[3]

3.2 Characterizing the Fixed Points

Next, our main goal is to obtain an RVU bound for $\max\{0, \mathrm{Reg}_{\Psi_i}^T\}$, but cast in terms of the player's strategies $(\boldsymbol{x}_i^{(t)})_{1 \leq t \leq T}$, as well as the utilities $(\boldsymbol{u}_i^{(t)})_{1 \leq t \leq T}$ observed by that player. In particular, in light of Propositions 3.1 and 3.2, the crux is to appropriately bound $\|\boldsymbol{\lambda}_i^{(t+1)} - \boldsymbol{\lambda}_i^{(t)}\|_{\boldsymbol{\lambda}_i^{(t)}, \infty}$ and $\sum_{\hat{\sigma} \in \Sigma_i^*} \|\boldsymbol{q}_{\hat{\sigma}}^{(t+1)} - \boldsymbol{q}_{\hat{\sigma}}^{(t)}\|_{\boldsymbol{q}_{\hat{\sigma}}^{(t)}, \infty}$ in terms of $\|\boldsymbol{x}_i^{(t+1)} - \boldsymbol{x}_i^{(t)}\|$—the deviation of the player's strategy at every time t. To do so, we prove the following key result.

Lemma 3.5. *Let $\boldsymbol{X}_i^{(t)} \in \mathbb{R}_{>0}^D$ be defined for every time $t \in \mathbb{N}$, for some $D \in \mathbb{N}$. Further, suppose that for every time $t \in \mathbb{N}$ and $\sigma \in \Sigma_i$,*

$$\boldsymbol{x}_i^{(t)}[\sigma] = \sum_{k=1}^{m} \frac{p_{\sigma,k}(\boldsymbol{X}_i^{(t)})}{q_{\sigma,k}(\boldsymbol{X}_i^{(t)})}, \tag{4}$$

for some multivariate polynomials $\{p_{\sigma,k}\}, \{q_{\sigma,k}\}$ with positive coefficients and maximum degree $\deg_i \in \mathbb{N}$. *If*

$$\max_{e \in [\![D]\!]} \left| 1 - \frac{\boldsymbol{X}_i^{(t+1)}[e]}{\boldsymbol{X}_i^{(t)}[e]} \right| \leq \frac{100}{256 \deg_i}, \tag{5}$$

it holds that

$$\|\boldsymbol{x}_i^{(t+1)} - \boldsymbol{x}_i^{(t)}\|_1 \leq (4\|\mathcal{Q}_i\|_1 \deg_i) \max_{e \in [\![D]\!]} \left| 1 - \frac{\boldsymbol{X}_i^{(t+1)}[e]}{\boldsymbol{X}_i^{(t)}[e]} \right|.$$

We recall that, based on Algorithm 1 (Line 3), the final strategy $\boldsymbol{x}_i^{(t)}$ is simply a fixed point of $\phi_i^{(t)} \in \Psi_i^{(t)}$, where $\phi_i^{(t)}$ is a function of $\boldsymbol{X}_i^{(t)} = (\boldsymbol{\lambda}_i^{(t)}, (\boldsymbol{q}_{\hat{\sigma}}^{(t)})_{\hat{\sigma} \in \Sigma_i^*})$. Equation (4) postulates that the fixed point is given by a rational function with positive coefficients. Taking a step back, let us

[3] Liu et al. [2022] very recently obtained near-optimal rates in zero-sum games, though with very different techniques.

clarify that assumption in the context of the no-swap-regret algorithm of Blum and Mansour [2007], a specific instance of Algorithm 1. In that algorithm, the fixed point is a stationary distribution of the underlying stochastic matrix $\boldsymbol{X}_i$; hence, Equation (4) is simply a consequence of the *Markov chain tree theorem* [Anantharam and Tsoucas, 1989], with degree in the order of the rank of the corresponding stochastic matrix.

While insisting on having positive coefficients in Lemma 3.5 may seem restrictive at first glance, in Propositions B.5 and B.6 (in Appendix B) we show that it comes without any loss under sequence-form vectors. We further remark that the degree of the rational function is a measure of the complexity of the fixed points, as it will be highlighted in Propositions 3.6 and 3.7 below. Finally, the property in (5) will be satisfied for our construction since the regret minimizers we employ guarantee *multiplicative stability* (shown in Lemmas B.2 and B.4).

We now establish that assumption (4) is satisfied for transformations in Ψ_i with only a moderate degree. First, as a warm-up, we consider fixed points associated with *coarse* trigger deviation functions $\tilde{\Psi}_i$.

Proposition 3.6. *Let $\phi_i^{(t)} \in \tilde{\Psi}_i$ be a transformation defined by $\boldsymbol{X}_i^{(t)} = (\boldsymbol{\lambda}_i^{(t)}, (\boldsymbol{q}_j^{(t)})_{j \in \mathcal{J}_i}) \in \mathbb{R}^D_{>0}$, for some $D \in \mathbb{N}$ and time $t \in \mathbb{N}$. The unique fixed point $\boldsymbol{x}_i^{(t)}$ of $\phi_i^{(t)}$ satisfies (4) with $\deg_i \leq 2\mathfrak{D}_i$.*

This property is established by leveraging the closed-form characterization for the fixed points associated with EFCCE given in [Anagnostides et al., 2022b]. Next, let us focus on the fixed points of trigger deviation functions. Unlike EFCCE, determining such fixed points requires computing stationary distributions of Markov chains along paths of the tree, commencing from the root and gradually making way towards the leaves [Farina et al., 2021]; this substantially complicates the analysis. Nevertheless, we leverage a refined characterization of the stationary distribution at every information set [Anagnostides et al., 2022b] to obtain the following.

Proposition 3.7. *Let $\phi_i^{(t)} \in \Psi_i$ be a transformation defined by $\boldsymbol{X}_i^{(t)} = (\boldsymbol{\lambda}_i^{(t)}, (\boldsymbol{q}_{\hat{\sigma}}^{(t)})_{\hat{\sigma} \in \Sigma_i^*}) \in \mathbb{R}^D_{>0}$, for some $D \in \mathbb{N}$ and time $t \in \mathbb{N}$. The (unique) fixed point $\boldsymbol{x}_i^{(t)}$ of $\phi_i^{(t)}$ satisfies (4) with $\deg_i \leq 2\mathfrak{D}_i|\mathcal{A}_i|$, where $|\mathcal{A}_i| \coloneqq \max_{j \in \mathcal{J}_i} |\mathcal{A}_j|$.*

In proof, we show that augmenting a "partial fixed point" at a new (successor) information set can only increase the degree of the rational function by an additive factor of $2|\mathcal{A}_i|$; Proposition 3.7 then follows by induction. It is crucial to note that using the Markov chain tree theorem directly at every information set would only give a bound on the degree that could be exponential in the description of the game. Next, we combine Proposition 3.7 with Lemma 3.5 to derive the following key inequality.

Lemma 3.8. *Consider any parameters $\eta \leq \frac{1}{256\|\mathcal{Q}_i\|_1 \deg_i}$ and $\eta_\triangle \leq \frac{1}{512|\Sigma_i| \deg_i}$, where $\deg_i \coloneqq 2|\mathcal{A}_i|\mathfrak{D}_i$. Then, for any time $t \in [\![T-1]\!]$,*

$$\|\boldsymbol{x}_i^{(t+1)} - \boldsymbol{x}_i^{(t)}\|_1 \leq 8\|\mathcal{Q}_i\|_1 |\mathcal{A}_i| \mathfrak{D}_i M(\boldsymbol{X}_i^{(t)}),$$

where $M(\boldsymbol{X}_i^{(t)})$ is defined as

$$\max \left\{ \max_{\hat{\sigma} \in \Sigma_i^*} \left| 1 - \frac{\boldsymbol{\lambda}_i^{(t+1)}[\hat{\sigma}]}{\boldsymbol{\lambda}_i^{(t)}[\hat{\sigma}]} \right|, \max_{\hat{\sigma} \in \Sigma_i^*} \max_{\sigma \in \Sigma_i} \left| 1 - \frac{\boldsymbol{q}_{\hat{\sigma}}^{(t+1)}[\sigma]}{\boldsymbol{q}_{\hat{\sigma}}^{(t)}[\sigma]} \right| \right\}.$$

3.3 Putting Everything Together

We can now combine Lemma 3.8 with Propositions 3.1 to 3.3, as well as some additional manipulations of the utilities $(\boldsymbol{U}_i^{(t)})_{1 \leq t \leq T}$ (Proposition 3.1) and $(\boldsymbol{u}_\triangle^{(t)})_{1 \leq t \leq T}$ (Proposition 3.2) to derive the following RVU bound.

Corollary 3.9. *Suppose that* $\eta \leq \frac{1}{2^{12}|\Sigma_i|^{1.5}\|\mathcal{Q}_i\|_1 \deg_i}$ *and* $\eta_\triangle = \frac{1}{2|\Sigma_i|}\eta$*, where* $\deg_i := 2|\mathcal{A}_i|\mathfrak{D}_i$*. For any* $T \geq 2$*,* $\max\{0, \mathrm{Reg}_{\Psi_i}^T\}$ *can be upper bounded by*

$$\frac{8|\Sigma_i|^2 \log T}{\eta} + 256\eta|\Sigma_i|^3 \sum_{t=1}^{T-1} \|\boldsymbol{u}_i^{(t+1)} - \boldsymbol{u}_i^{(t)}\|_\infty^2 - \frac{1}{2^{15}\eta \deg_i^2 \|\mathcal{Q}_i\|_1^2} \sum_{t=1}^{T-1} \|\boldsymbol{x}_i^{(t+1)} - \boldsymbol{x}_i^{(t)}\|_1^2.$$

Given that the RVU bound in Corollary 3.9 has been obtained for $\max\{0, \mathrm{Reg}_{\Psi_i}^T\}$, a nonnegative measure of regret, we can show that the second-order path lengths of the dynamics are bounded by $O(\log T)$; that is,

$$\sum_{t=1}^{T-1} \sum_{i=1}^{n} \|\boldsymbol{x}_i^{(t+1)} - \boldsymbol{x}_i^{(t)}\|_1^2 = O(\log T).$$

This step is formalized in Theorem B.12 (in Appendix B), and follows the technique of Anagnostides et al. [2022c], leading to our main result; below we use the notation $|\Sigma| := \max_{i \in [\![n]\!]} |\Sigma_i|$, and similarly for the other symbols (namely, $\|\mathcal{Q}\|_1, |\mathcal{A}|$ and $\mathfrak{D}$).

Theorem 3.10. *If all players employ Algorithm 1 instantiated with* `LRL-OFTRL` *for all local regret minimizers,* $\mathfrak{R}_\triangle$ *and* $\{\mathfrak{R}_{\hat{\sigma}}\}_{\hat{\sigma}}$*, the trigger regret of each player* $i \in [\![n]\!]$ *after* T *repetitions will be bounded as*

$$\mathrm{Reg}_{\Psi_i}^T \leq Cn|\Sigma|^{3.5}\|\mathcal{Q}\|_1|\mathcal{Z}||\mathcal{A}|\mathfrak{D}\log T,$$

for a universal constant $C > 0$*.*

For EFCCE, in accordance to Proposition 3.6, we obtain a slightly improved regret bound (Corollary B.14 in Appendix B).

4 Experimental Results

Finally, in this section we experimentally verify our theoretical results on several common benchmark extensive-form games: (i) 3-player *Kuhn poker* [Kuhn, 1953]; (ii) 2-player *Goofspiel* [Ross, 1971]; and (iii) 2-player *Sheriff* [Farina et al., 2019c]. All of these are general-sum games. A detailed description of the game instances we use is included in Appendix C.

In accordance to Theorem 3.10, we instantiate each local regret minimizer using `LRL-OFTRL`, and all players use the same learning algorithm. For simplicity we use the same learning rate $\eta > 0$ for all the local regret minimizers, which is treated as a hyperparameter in order to obtain better empirical performance. In particular, after a very mild tuning process, we chose $\eta = 1$ for all our experiments. We compare the performance of our algorithm with that of two other popular regret minimizers: 1) CFR with regret matching (RM) [Zinkevich et al., 2007], meaning that every local regret minimizer $\mathfrak{R}_{\hat{\sigma}}$ uses CFR (with RM) and $\mathfrak{R}_\triangle$ (which is an algorithm for the simplex) also uses RM; and 2) CFR$^+$ with RM$^+$ [Tammelin et al., 2015]. We did not employ alternation or linear averaging, two popular tricks that accelerate convergence in zero-sum games, as it is not known if those techniques retain convergence in our setting.

Our findings are illustrated in Figure 1. As predicted by our theory (Theorem 3.10), the trigger regret of all players appears to grow as $O(\log T)$ (the x-axis is logarithmic), implying convergence to the set of EFCE with a rate of $\frac{\log T}{T}$. In contrast, although the trigger regret experienced by the other regret minimizers is sometimes smaller compared to our algorithm, their asymptotic growth appears to exhibit an unfavorable exponential increase, meaning that their trigger regret grows as $\omega(\log T)$, with the exception of 3-player Kuhn poker. In fact, for Kuhn poker we see that the learning dynamics actually converge after only a few iterations, but this is not typical behavior in general-sum games. Indeed, for the other two games in Figure 1 we did not observe convergence. We also obtained qualitatively similar regret bounds for *coarse* trigger regret—associated with EFCCE.

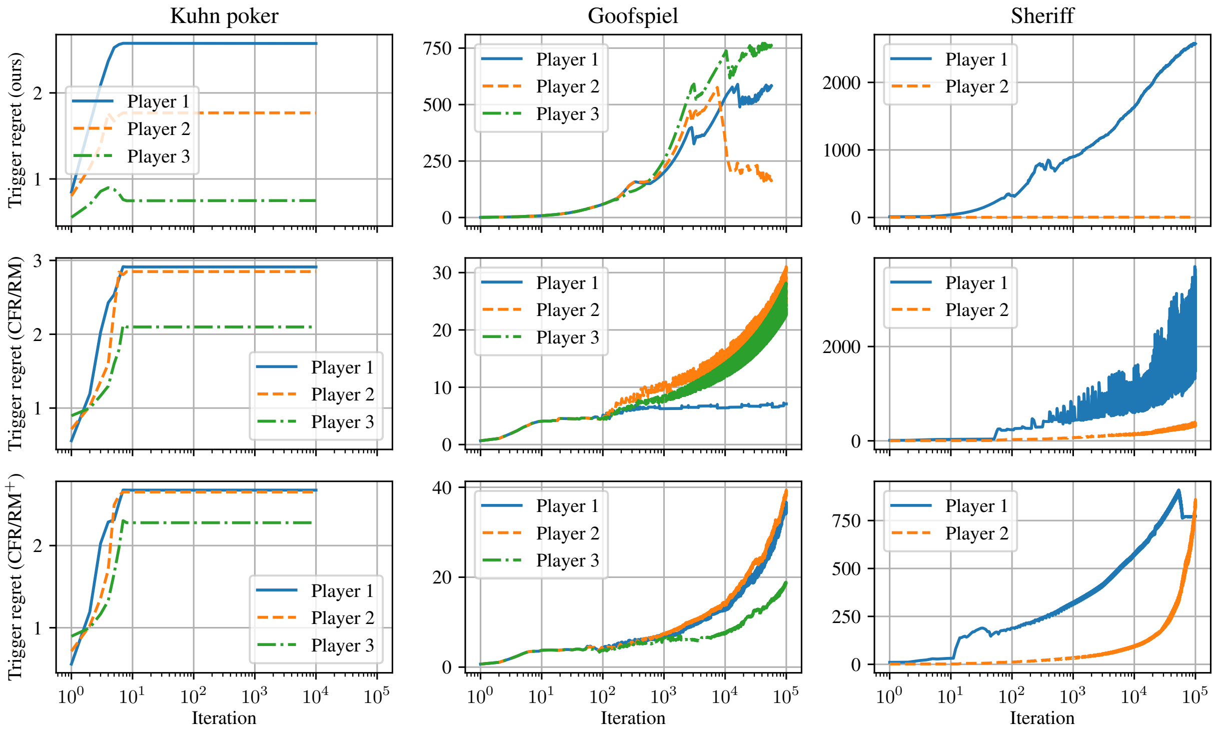

Figure 1: Trigger regret of each player on (i) Kuhn poker (left); (ii) Goofspiel (center); and (iii) Sheriff (right). Every row corresponds to a different algorithm, starting from ours in the first one. The x-axis indicates the iteration, while the y-axis indicates the corresponding trigger regret for each player. We emphasize that the x-axis is logarithmic.

5 Conclusions and Future Research

In this paper, we established the first near-optimal $\frac{\log T}{T}$ rates of convergence to extensive-form correlated equilibria, thereby extending recent work from normal-form games to the substantially more complex class of imperfect-information extensive-form games. Our approach for obtaining near-optimal Φ-regret guarantees can be in fact further extended beyond extensive-form games, as long as the fixed points admit the characterization imposed by Lemma 3.5. Our techniques also have an independent interest in deriving near-optimal rates using the regret-decomposition approach, a question that previously remained elusive.

There are still many interesting avenues for future research. While our trigger-regret bounds are near-optimal in terms of the dependence on T (Theorem 3.10), the dependence on the parameters of the game can likely be improved. Establishing near-optimal trigger regret under dynamics that do not employ logarithmic regularization, such as optimistic hedge, is another challenging open problem; it is plausible that the techniques of Anagnostides et al. [2022a] could be useful in that direction.

References

Geoffrey J Gordon, Amy Greenwald, and Casey Marks. No-regret learning in convex games. In *Proceedings of the 25th international conference on Machine learning*, pages 360–367. ACM, 2008.

Elad Hazan and Satyen Kale. Computational equivalence of fixed points and no regret algorithms, and convergence to equilibria. In *Advances in Neural Information Processing Systems 20, Proceedings of the Twenty-First Annual Conference on Neural Information Processing Systems, 2007*, pages 625–632. Curran Associates, Inc., 2007.

Alexander Rakhlin, Karthik Sridharan, and Ambuj Tewari. Online learning: Beyond regret. In Sham M. Kakade and Ulrike von Luxburg, editors, *COLT 2011 - The 24th Annual Conference on Learning Theory, 2011*, volume 19 of *JMLR Proceedings*, pages 559–594. JMLR.org, 2011.

Gilles Stoltz and Gábor Lugosi. Learning correlated equilibria in games with compact sets of strategies. *Games Econ. Behav.*, 59(1):187–208, 2007.

Amy Greenwald and Amir Jafari. A general class of no-regret learning algorithms and game-theoretic equilibria. In *Computational Learning Theory and Kernel Machines, COLT/Kernel 2003*, volume 2777 of *Lecture Notes in Computer Science*, pages 2–12. Springer, 2003.

Dean Foster and Rakesh Vohra. Calibrated learning and correlated equilibrium. *Games and Economic Behavior*, 21:40–55, 1997.

Sergiu Hart and Andreu Mas-Colell. A simple adaptive procedure leading to correlated equilibrium. *Econometrica*, 68:1127–1150, 2000.

Constantinos Daskalakis, Alan Deckelbaum, and Anthony Kim. Near-optimal no-regret algorithms for zero-sum games. In *Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2011.

Alexander Rakhlin and Karthik Sridharan. Online learning with predictable sequences. In *Conference on Learning Theory*, pages 993–1019, 2013a.

Alexander Rakhlin and Karthik Sridharan. Optimization, learning, and games with predictable sequences. In *Advances in Neural Information Processing Systems*, pages 3066–3074, 2013b.

Vasilis Syrgkanis, Alekh Agarwal, Haipeng Luo, and Robert E Schapire. Fast convergence of regularized learning in games. In *Advances in Neural Information Processing Systems*, pages 2989–2997, 2015.

Dylan J. Foster, Zhiyuan Li, Thodoris Lykouris, Karthik Sridharan, and Éva Tardos. Learning in games: Robustness of fast convergence. In *Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems 2016*, pages 4727–4735, 2016.

Chen-Yu Wei and Haipeng Luo. More adaptive algorithms for adversarial bandits. In *Conference On Learning Theory, COLT 2018*, volume 75 of *Proceedings of Machine Learning Research*, pages 1263–1291. PMLR, 2018.

Xi Chen and Binghui Peng. Hedging in games: Faster convergence of external and swap regrets. In *Proceedings of the Annual Conference on Neural Information Processing Systems (NeurIPS)*, 2020.

Yu-Guan Hsieh, Kimon Antonakopoulos, and Panayotis Mertikopoulos. Adaptive learning in continuous games: Optimal regret bounds and convergence to nash equilibrium. In Mikhail Belkin and Samory Kpotufe, editors, *Conference on Learning Theory, COLT 2021*, volume 134 of *Proceedings of Machine Learning Research*, pages 2388–2422. PMLR, 2021.

Constantinos Daskalakis, Maxwell Fishelson, and Noah Golowich. Near-optimal no-regret learning in general games. In *Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021*, pages 27604–27616, 2021.

Constantinos Daskalakis and Noah Golowich. Fast rates for nonparametric online learning: from realizability to learning in games. In *STOC '22: 54th Annual ACM SIGACT Symposium on Theory of Computing*, pages 846–859. ACM, 2022.

Ioannis Anagnostides, Constantinos Daskalakis, Gabriele Farina, Maxwell Fishelson, Noah Golowich, and Tuomas Sandholm. Near-optimal no-regret learning for correlated equilibria in multi-player general-sum games. In *STOC '22: 54th Annual ACM SIGACT Symposium on Theory of Computing*, pages 736–749. ACM, 2022a.

Georgios Piliouras, Ryann Sim, and Stratis Skoulakis. Optimal no-regret learning in general games: Bounded regret with unbounded step-sizes via clairvoyant mwu. *arXiv preprint arXiv:2111.14737*, 2021.

Bernhard Von Stengel and Françoise Forges. Extensive-form correlated equilibrium: Definition and computational complexity. *Mathematics of Operations Research*, 33(4):1002–1022, 2008.

Andrea Celli, Alberto Marchesi, Gabriele Farina, and Nicola Gatti. No-regret learning dynamics for extensive-form correlated equilibrium. In *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020*, 2020.

Dustin Morrill, Ryan D'Orazio, Marc Lanctot, James R. Wright, Michael Bowling, and Amy R. Greenwald. Efficient deviation types and learning for hindsight rationality in extensive-form games. In *Proceedings of the 38th International Conference on Machine Learning, ICML 2021*, volume 139 of *Proceedings of Machine Learning Research*, pages 7818–7828. PMLR, 2021a.

Ioannis Anagnostides, Gabriele Farina, Christian Kroer, Andrea Celli, and Tuomas Sandholm. Faster no-regret learning dynamics for extensive-form correlated and coarse correlated equilibria. In *EC '22: The 23rd ACM Conference on Economics and Computation, 2022*, pages 915–916. ACM, 2022b.

Dustin Morrill, Ryan D'Orazio, Reca Sarfati, Marc Lanctot, James R. Wright, Amy R. Greenwald, and Michael Bowling. Hindsight and sequential rationality of correlated play. In *Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021*, pages 5584–5594. AAAI Press, 2021b.

Yu Bai, Chi Jin, Song Mei, Ziang Song, and Tiancheng Yu. Efficient Φ-regret minimization in extensive-form games via online mirror descent. *CoRR*, abs/2205.15294, 2022a.

Ziang Song, Song Mei, and Yu Bai. Sample-efficient learning of correlated equilibria in extensive-form games. *CoRR*, abs/2205.07223, 2022.

Gabriele Farina, Andrea Celli, Alberto Marchesi, and Nicola Gatti. Simple uncoupled no-regret learning dynamics for extensive-form correlated equilibrium. *CoRR*, abs/2104.01520, 2021.

Yu Bai, Chi Jin, Song Mei, and Tiancheng Yu. Near-optimal learning of extensive-form games with imperfect information. In *International Conference on Machine Learning, ICML 2022*, volume 162 of *Proceedings of Machine Learning Research*, pages 1337–1382. PMLR, 2022b.

Gabriele Farina, Ioannis Anagnostides, Haipeng Luo, Chung-Wei Lee, Christian Kroer, and Tuomas Sandholm. Near-optimal no-regret learning for general convex games. *CoRR*, abs/2206.08742, 2022.

Martin Zinkevich, Michael Bowling, Michael Johanson, and Carmelo Piccione. Regret minimization in games with incomplete information. In *Proceedings of the Annual Conference on Neural Information Processing Systems (NIPS)*, 2007.

Ioannis Anagnostides, Gabriele Farina, Christian Kroer, Chung-Wei Lee, Haipeng Luo, and Tuomas Sandholm. Uncoupled learning dynamics with $O(\log T)$ swap regret in multiplayer games. *CoRR*, abs/2204.11417, 2022c.

Gabriele Farina, Tommaso Bianchi, and Tuomas Sandholm. Coarse correlation in extensive-form games. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 1934–1941, 2020.

Casey Alvin Marks. *No-regret learning and game-theoretic equilibria.* Brown University, 2008.

Georgios Piliouras, Mark Rowland, Shayegan Omidshafiei, Romuald Elie, Daniel Hennes, Jerome T. Connor, and Karl Tuyls. Evolutionary dynamics and phi-regret minimization in games. *J. Artif. Intell. Res.*, 74:1125–1158, 2022.

H. Moulin and J.-P. Vial. Strategically zero-sum games: The class of games whose completely mixed equilibria cannot be improved upon. *International Journal of Game Theory*, 7(3-4):201–221, 1978.

Robert Aumann. Subjectivity and correlation in randomized strategies. *Journal of Mathematical Economics*, 1:67–96, 1974.

Eddie Dekel and Drew Fudenberg. Rational behavior with payoff uncertainty. *Journal of Economic Theory*, 52(2):243–267, 1990.

Yannick Viossat and Andriy Zapechelnyuk. No-regret dynamics and fictitious play. *J. Econ. Theory*, 148(2):825–842, 2013.

Angeliki Giannou, Emmanouil-Vasileios Vlatakis-Gkaragkounis, and Panayotis Mertikopoulos. Survival of the strictest: Stable and unstable equilibria under regularized learning with partial information. In Mikhail Belkin and Samory Kpotufe, editors, *Conference on Learning Theory, COLT 2021*, volume 134 of *Proceedings of Machine Learning Research*, pages 2147–2148. PMLR, 2021.

Hugh Zhang. A simple adaptive procedure converging to forgiving correlated equilibria. *CoRR*, abs/2207.06548, 2022.

Miroslav Dudík and Geoffrey J. Gordon. A sampling-based approach to computing equilibria in succinct extensive-form games. In Jeff A. Bilmes and Andrew Y. Ng, editors, *UAI 2009, Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence, 2009*, pages 151–160. AUAI Press, 2009.

Wan Huang and Bernhard von Stengel. Computing an extensive-form correlated equilibrium in polynomial time. In *Internet and Network Economics, 4th International Workshop, WINE 2008*, volume 5385 of *Lecture Notes in Computer Science*, pages 506–513. Springer, 2008.

Nicolo Cesa-Bianchi and Gabor Lugosi. *Prediction, learning, and games.* Cambridge University Press, 2006.

Yu-Guan Hsieh, Kimon Antonakopoulos, Volkan Cevher, and Panayotis Mertikopoulos. No-regret learning in games with noisy feedback: Faster rates and adaptivity via learning rate separation. *CoRR*, abs/2206.06015, 2022.

Liad Erez, Tal Lancewicki, Uri Sherman, Tomer Koren, and Yishay Mansour. Regret minimization and convergence to equilibria in general-sum markov games. *CoRR*, abs/2207.14211, 2022.

Runyu Zhang, Qinghua Liu, Huan Wang, Caiming Xiong, Na Li, and Yu Bai. Policy optimization for markov games: Unified framework and faster convergence. *CoRR*, abs/2206.02640, 2022.

Kevin Leyton-Brown and Yoav Shoham. *Essentials of Game Theory: A Concise Multidisciplinary Introduction.* Synthesis Lectures on Artificial Intelligence and Machine Learning. Morgan & Claypool Publishers, 2008.

Gabriele Farina, Christian Kroer, and Tuomas Sandholm. Regret circuits: Composability of regret minimizers. In *International Conference on Machine Learning*, pages 1863–1872, 2019a.

Gabriele Farina, Christian Kroer, Noam Brown, and Tuomas Sandholm. Stable-predictive optimistic counterfactual regret minimization. In *International Conference on Machine Learning (ICML)*, 2019b.

Mingyang Liu, Asuman E. Ozdaglar, Tiancheng Yu, and Kaiqing Zhang. The power of regularization in solving extensive-form games. *CoRR*, abs/2206.09495, 2022.

Avrim Blum and Yishay Mansour. From external to internal regret. *J. Mach. Learn. Res.*, 8: 1307–1324, 2007.

V. Anantharam and P. Tsoucas. A proof of the markov chain tree theorem. *Statistics & Probability Letters*, 8(2):189–192, 1989.

H. W. Kuhn. Extensive games and the problem of information. In H. W. Kuhn and A. W. Tucker, editors, *Contributions to the Theory of Games*, volume 2 of *Annals of Mathematics Studies, 28*, pages 193–216. Princeton University Press, Princeton, NJ, 1953.

Sheldon M Ross. Goofspiel—the game of pure strategy. *Journal of Applied Probability*, 8(3): 621–625, 1971.

Gabriele Farina, Chun Kai Ling, Fei Fang, and Tuomas Sandholm. Correlation in extensive-form games: Saddle-point formulation and benchmarks. In *Conference on Neural Information Processing Systems (NeurIPS)*, 2019c.

Oskari Tammelin, Neil Burch, Michael Johanson, and Michael Bowling. Solving heads-up limit Texas hold'em. In *Proceedings of the 24th International Joint Conference on Artificial Intelligence (IJCAI)*, 2015.

A Additional Preliminaries

In this section, we provide some additional background on extensive-form games and (coarse) trigger deviation functions.

An illustrative example First, to clarify some of the concepts we introduced earlier in Section 2, we consider the simple two-player EFG of Figure 2. White round nodes correspond to player 1, while black round nodes to player 2. We use square nodes to represent terminal nodes (or leaves). As illustrated in Figure 2, player 1 has two information sets, denoted by $\mathcal{J}_1 := \{\mathrm{A}, \mathrm{B}\}$, each containing two nodes. Further, the set of sequences of player 1 can be represented as $\Sigma_1 := \{\emptyset, \mathtt{1}, \mathtt{2}, \mathtt{3}, \mathtt{4}\}$; here, we omitted specifying the corresponding information set since we use different symbols for actions belonging to different information sets.

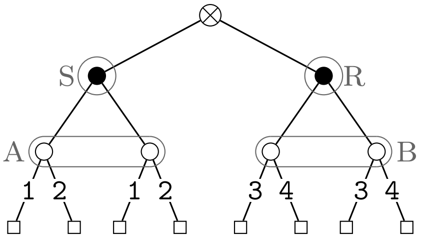

Figure 2: Example of a two-player EFG.

Before we proceed, let us clarify some notation that will be useful in the sequel. For any pair of sequences $\sigma, \sigma' \in \Sigma_i^*$, with $\sigma = (j, a)$ and $\sigma' = (j', a')$, we write $\sigma \prec \sigma'$ if the sequence of actions encountered from the root of the tree to any node in j' includes selecting action a at some node from information set j. Further, by convention, we let $\emptyset \prec \sigma$ for any $\sigma \in \Sigma_i^*$. We will also write $\sigma \succeq j$ if sequence σ must pass from some node in j.

Trigger deviation functions It will be convenient to represent trigger deviation functions, in the sense of Definition 2.2, as follows.

Definition A.1 ([Farina et al., 2021])**.** Let $\hat{\sigma} = (j, a) \in \Sigma_i^*$, and $\boldsymbol{q} \in \mathcal{Q}_j$. We let $\mathbf{M}_{\hat{\sigma} \to \boldsymbol{q}} \in \mathbb{R}^{|\Sigma_i| \times |\Sigma_i|}$ be a matrix, so that for any $\sigma_r, \sigma_c \in \Sigma_i$,

$$\mathbf{M}_{\hat{\sigma} \to \boldsymbol{q}} = \begin{cases} 1 & \text{if } \sigma_c \not\succeq \hat{\sigma} \text{ and } \sigma_r = \sigma_c; \\ \boldsymbol{q}[\sigma_r] & \text{if } \sigma_c = \hat{\sigma} \text{ and } \sigma_r \succeq j; \text{ and} \\ 0 & \text{otherwise.} \end{cases}$$

We will let $\phi_{\hat{\sigma} \to \boldsymbol{q}}$ denote the linear function $\boldsymbol{x} \mapsto \mathbf{M}_{\hat{\sigma} \to \boldsymbol{q}} \boldsymbol{x}$, for some $\boldsymbol{q} \in \mathcal{Q}_j$. It is immediate to verify that for any $\hat{\sigma} = (j, a) \in \Sigma_i^*$ and $\boldsymbol{q} \in \mathcal{Q}_j$, $\phi_{\hat{\sigma} \to \boldsymbol{q}}$ is a trigger deviation function in the sense of Definition 2.2.

To clarify Definition A.1, below we give two examples for the EFG of Figure 2. If $\boldsymbol{q} = (\frac{1}{2}, \frac{1}{2}) \in \Delta^2$, then

$$\mathbf{M}_{1\to\boldsymbol{q}} = \begin{array}{cccccc} & \varnothing & 1 & 2 & 3 & 4 \\ \varnothing & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & \mathbf{1/2} & 0 & 0 & 0 \\ 2 & 0 & \mathbf{1/2} & 1 & 0 & 0 \\ 3 & 0 & 0 & 0 & 1 & 0 \\ 4 & 0 & 0 & 0 & 0 & 1 \end{array}, \quad \mathbf{M}_{3\to\boldsymbol{q}} = \begin{array}{cccccc} & \varnothing & 1 & 2 & 3 & 4 \\ \varnothing & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 2 & 0 & 0 & 1 & 0 & 0 \\ 3 & 0 & 0 & 0 & \mathbf{1/2} & 0 \\ 4 & 0 & 0 & 0 & \mathbf{1/2} & 1 \end{array}.$$

The following characterization can be readily extracted from [Farina et al., 2021].

Claim A.2. *Every transformation $\phi_i \in \Psi_i$ can be expressed as $\sum_{\hat{\sigma} \in \Sigma_i^*} \boldsymbol{\lambda}_i[\hat{\sigma}] \phi_{\hat{\sigma} \to \boldsymbol{q}_{\hat{\sigma}}}$, where $\boldsymbol{\lambda}_i \in \Delta(\Sigma_i^*)$ and $\boldsymbol{q}_{\hat{\sigma}} \in \mathcal{Q}_j$ for $\hat{\sigma} = (j, a) \in \Sigma_i^*$.*

Coarse trigger deviation functions Analogously, *coarse* trigger deviation functions can be represented as follows.

Definition A.3 ([Anagnostides et al., 2022b]). Let $j \in \mathcal{J}_i$ and $\boldsymbol{q} \in \mathcal{Q}_j$. We let $\mathbf{M}_{j \to \boldsymbol{q}} \in \mathbb{R}^{|\Sigma_i| \times |\Sigma_i|}$ be a matrix, so that for any $\sigma_r, \sigma_c \in \Sigma_i$,

$$\mathbf{M}_{j \to \boldsymbol{q}} = \begin{cases} 1 & \text{if } \sigma_c \not\succeq j \text{ and } \sigma_r = \sigma_c; \\ \boldsymbol{q}[\sigma_r] & \text{if } \sigma_c = \sigma_j \text{ and } \sigma_r \succeq j; \text{ and} \\ 0 & \text{otherwise.} \end{cases}$$

Unlike trigger deviations, which are "triggered" by a sequence, we point out that *coarse* trigger deviations are "triggered" by an information set; see [Farina et al., 2020] for a more detailed discussion on this point.

Returning to the example of Figure 2, and letting again $\boldsymbol{q} = (\frac{1}{2}, \frac{1}{2}) \in \Delta^2$,

$$\mathbf{M}_{\mathrm{A}\to\boldsymbol{q}} = \begin{array}{cccccc} & \varnothing & 1 & 2 & 3 & 4 \\ \varnothing & 1 & 0 & 0 & 0 & 0 \\ 1 & \mathbf{1/2} & 0 & 0 & 0 & 0 \\ 2 & \mathbf{1/2} & 0 & 0 & 0 & 0 \\ 3 & 0 & 0 & 0 & 1 & 0 \\ 4 & 0 & 0 & 0 & 0 & 1 \end{array}, \quad \mathbf{M}_{\mathrm{C}\to\boldsymbol{q}} = \begin{array}{cccccc} & \varnothing & 1 & 2 & 3 & 4 \\ \varnothing & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 2 & 0 & 0 & 1 & 0 & 0 \\ 3 & \mathbf{1/2} & 0 & 0 & 0 & 0 \\ 4 & \mathbf{1/2} & 0 & 0 & 0 & 0 \end{array}.$$

Analogously to Claim A.2, one can show the following characterization.

Claim A.4. *Every transformation $\phi_i \in \tilde{\Psi}_i$ can be expressed as $\sum_{j \in \mathcal{J}_i} \boldsymbol{\lambda}_i[j] \phi_{j \to \boldsymbol{q}_j}$, where $\boldsymbol{\lambda}_i \in \Delta(\mathcal{J}_i)$ and $\boldsymbol{q}_j \in \mathcal{Q}_j$.*

The connection between coarse trigger deviation functions and EFCCE is illuminated in the following fact.

Theorem A.5 ([Anagnostides et al., 2022b]). *If each player i incurs coarse trigger regret $\mathrm{Reg}^T_{\tilde{\Psi}_i}$ after T repetitions of the game, the average product distribution of play is a $\frac{1}{T} \max_{i \in [\![n]\!]} \mathrm{Reg}^T_{\tilde{\Psi}_i}$-approximate EFCCE.*

B Omitted Proofs

In this section, we provide all the omitted proofs from the main body (Section 3). For the convenience of the reader, we restate each claim before proceedings with its proof.

B.1 RVU Bounds for the Set of Deviations

Let us fix a player $i \in [\![n]\!]$. First, we prove Proposition 3.1. To this end, let us provide some auxiliary claims. Recall that, for each $\hat{\sigma} = (j, a) \in \Sigma_i^*$, $\mathfrak{R}_{\hat{\sigma}}$ receives at every time t the utility $\boldsymbol{U}_i^{(t)} \coloneqq \boldsymbol{u}_i^{(t)} \otimes \boldsymbol{x}_i^{(t)}$, and the next strategy is computed via `LRL-OFTRL` [Farina et al., 2022]; namely, we first compute $\tilde{\boldsymbol{q}}_{\hat{\sigma}}^{(t)} = (\lambda_{\hat{\sigma}}^{(t)}, \boldsymbol{y}_{\hat{\sigma}}^{(t)}) \in \tilde{\mathcal{Q}}_j$, for a time $t \in \mathbb{N}$, as

$$\operatorname*{arg\,max}_{\tilde{\boldsymbol{q}}_{\hat{\sigma}} \in \tilde{\mathcal{Q}}_j} \left\{ \eta \left\langle \boldsymbol{S}_{\hat{\sigma}}^{(t-1)}, \tilde{\boldsymbol{q}}_{\hat{\sigma}} \right\rangle + \sum_{e \in \Sigma_j \cup \{0\}} \log \tilde{\boldsymbol{q}}_{\hat{\sigma}}[e] \right\}, \tag{6}$$

where,

(i) $\tilde{\mathcal{Q}}_j \coloneqq \{(\lambda_{\hat{\sigma}}, \boldsymbol{y}_{\hat{\sigma}}) : \lambda_{\hat{\sigma}} \in [0,1], \boldsymbol{y}_{\hat{\sigma}} \in \lambda_{\hat{\sigma}} \mathcal{Q}_j\}$;

(ii) $\tilde{\boldsymbol{U}}_{\hat{\sigma}}^{(t)} \coloneqq (-\langle \boldsymbol{q}_{\hat{\sigma}}^{(t)}, \boldsymbol{U}_{\hat{\sigma}}^{(t)} \rangle, \boldsymbol{U}_{\hat{\sigma}}^{(t)})$, where in turn $\boldsymbol{U}_{\hat{\sigma}}^{(t)}$ is the component of $\boldsymbol{U}_i^{(t)}$ that corresponds to sequence $\hat{\sigma}$;

(iii) $\boldsymbol{S}_{\hat{\sigma}}^{(t-1)} \coloneqq \tilde{\boldsymbol{U}}_{\hat{\sigma}}^{(t-1)} + \sum_{\tau=1}^{t-1} \tilde{\boldsymbol{U}}_{\hat{\sigma}}^{(\tau)}$; and

(iv) $\eta > 0$ is the learning rate—common among all $\mathfrak{R}_{\hat{\sigma}}$.

Finally, having determined $\tilde{\boldsymbol{q}}_{\hat{\sigma}}^{(t)} = (\lambda_{\hat{\sigma}}^{(t)}, \boldsymbol{y}_{\hat{\sigma}}^{(t)})$, we compute $\boldsymbol{q}_{\hat{\sigma}}^{(t)} \coloneqq \frac{\boldsymbol{y}_{\hat{\sigma}}^{(t)}}{\lambda_{\hat{\sigma}}^{(t)}} \in \mathcal{Q}_j$. In turn, this gives the next strategy of $\mathfrak{R}_{\hat{\sigma}}$ as $\boldsymbol{X}_{\hat{\sigma}}^{(t)} \coloneqq \mathbf{M}_{\hat{\sigma} \to \boldsymbol{q}_{\hat{\sigma}}^{(t)}}$ (recall Definition A.1). It is evident that the regret minimization problem faced by each $\mathfrak{R}_{\hat{\sigma}}$ is equivalent to minimizing regret over $\mathcal{Q}_j$, since only the components of $\boldsymbol{X}_{\hat{\sigma}}^{(t)}$ that correspond to $\boldsymbol{q}_{\hat{\sigma}}^{(t)}$ cumulate regret (the rest are constant), leading to the regret bound below. We note that all the subsequent analysis operates under the tacit premise that each local regret minimizer is updated via `LRL-OFTRL`, without explicitly mentioned in the statements in order to lighten the exposition.

Proposition B.1. *For any learning rate* $\eta \leq \frac{1}{256 \|\mathcal{Q}_i\|_1}$ *and* $T \geq 2$, $\max\{0, \mathrm{Reg}_{\hat{\sigma}}^T\}$ *can be upper bounded by*

$$2 \frac{|\Sigma_i| \log T}{\eta} + 16 \eta \|\mathcal{Q}_i\|^2 \sum_{t=1}^{T-1} \|\boldsymbol{U}_i^{(t+1)} - \boldsymbol{U}_i^{(t)}\|_\infty^2 - \frac{1}{32\eta} \sum_{t=1}^{T-1} \left\| \begin{pmatrix} \lambda_{\hat{\sigma}}^{(t+1)} \\ \boldsymbol{y}_{\hat{\sigma}}^{(t+1} \end{pmatrix} - \begin{pmatrix} \lambda_{\hat{\sigma}}^{(t)} \\ \boldsymbol{y}_{\hat{\sigma}}^{(t} \end{pmatrix} \right\|_t^2 .$$

Proof. This regret bound is an immediate implication of [Farina et al., 2022, Proposition 2 and Corollary 1]. More precisely, we note that the regret bound in [Farina et al., 2022] applies if $\|\boldsymbol{U}_i^{(t)}\|_\infty \leq \frac{1}{\|\mathcal{Q}_i\|_1}$, for any $t \in \mathbb{N}$. That assumption can be met by rescaling the learning rate by a factor of $\frac{1}{\|\mathcal{Q}_i\|_1}$ since in our setting it holds that $\|\boldsymbol{U}_i^{(t)}\|_\infty \leq 1$; the latter follows from the definition of $\boldsymbol{U}_i^{(t)} \coloneqq \boldsymbol{u}_i^{(t)} \otimes \boldsymbol{x}_i^{(t)}$ (Line 6), and the fact that $\|\boldsymbol{u}_i^{(t)}\|_\infty \leq 1$ (by assumption) and $\|\boldsymbol{x}_i^{(t)}\|_\infty \leq 1$ (since $\mathcal{Q}_i \subseteq [0,1]^{|\Sigma_i|}$). □

In Proposition B.1 we used the shorthand notation

$$\left\| \begin{pmatrix} \lambda_{\hat{\sigma}}^{(t+1)} \\ \boldsymbol{y}_{\hat{\sigma}}^{(t+1} \end{pmatrix} - \begin{pmatrix} \lambda_{\hat{\sigma}}^{(t)} \\ \boldsymbol{y}_{\hat{\sigma}}^{(t} \end{pmatrix} \right\|_t^2 := \left\| \begin{pmatrix} \lambda_{\hat{\sigma}}^{(t+1)} \\ \boldsymbol{y}_{\hat{\sigma}}^{(t+1} \end{pmatrix} - \begin{pmatrix} \lambda_{\hat{\sigma}}^{(t)} \\ \boldsymbol{y}_{\hat{\sigma}}^{(t} \end{pmatrix} \right\|_{(\lambda_{\hat{\sigma}}^{(t)}, \boldsymbol{y}_{\hat{\sigma}}^{(t)})}^2,$$

where for a vector $\tilde{\boldsymbol{w}} \in \mathbb{R}^{d+1}$ and $\tilde{\boldsymbol{x}} \in \mathbb{R}_{>0}^{d+1}$, we used the notation

$$\|\tilde{\boldsymbol{w}}\|_{\tilde{\boldsymbol{x}}} := \sqrt{\sum_{e \in [\![d+1]\!]} \left(\frac{\tilde{\boldsymbol{w}}[e]}{\tilde{\boldsymbol{x}}[e]} \right)^2}$$

for the local norm induced by $\tilde{\boldsymbol{x}}$. Further, we will also use the notation

$$\|\tilde{\boldsymbol{w}}\|_{\tilde{\boldsymbol{x}},\infty} := \max_{e \in [\![d+1]\!]} \left| \frac{\tilde{\boldsymbol{w}}[e]}{\tilde{\boldsymbol{x}}[e]} \right|.$$

Lemma B.2. *For any sequence $\hat{\sigma} = (j, a) \in \Sigma_i^*$, learning rate $\eta \leq \frac{1}{50\|\mathcal{Q}_i\|_1}$ and time $t \in [\![T-1]\!]$,*

$$\max_{\sigma \in \Sigma_j} \left| 1 - \frac{\boldsymbol{q}_{\hat{\sigma}}^{(t+1)}[\sigma]}{\boldsymbol{q}_{\hat{\sigma}}^{(t)}[\sigma]} \right| \leq 4 \left\| \begin{pmatrix} \lambda_{\hat{\sigma}}^{(t+1)} \\ \boldsymbol{y}_{\hat{\sigma}}^{(t+1)} \end{pmatrix} - \begin{pmatrix} \lambda_{\hat{\sigma}}^{(t)} \\ \boldsymbol{y}_{\hat{\sigma}}^{(t)} \end{pmatrix} \right\|_{t,\infty} \leq 100\eta\|\mathcal{Q}_i\|_1.$$

Proof. We will need the following stability bound, extracted from [Farina et al., 2022, Proposition 3].

Lemma B.3 ([Farina et al., 2022])**.** *For any sequence $\hat{\sigma} = (j, a) \in \Sigma_i^*$, time $t \in [\![T-1]\!]$ and learning rate $\eta \leq \frac{1}{50\|\mathcal{Q}_i\|_1}$,*

$$\left\| \begin{pmatrix} \lambda_{\hat{\sigma}}^{(t+1)} \\ \boldsymbol{y}_{\hat{\sigma}}^{(t+1)} \end{pmatrix} - \begin{pmatrix} \lambda_{\hat{\sigma}}^{(t)} \\ \boldsymbol{y}_{\hat{\sigma}}^{(t)} \end{pmatrix} \right\|_{t,\infty} \leq 22\eta\|\mathcal{Q}_i\|_1.$$

Now let us fix a time $t \in [\![T-1]\!]$. For convenience, we introduce the notation

$$\mu^{(t)} := \left\| \begin{pmatrix} \lambda_{\hat{\sigma}}^{(t+1)} \\ \boldsymbol{y}_{\hat{\sigma}}^{(t+1)} \end{pmatrix} - \begin{pmatrix} \lambda_{\hat{\sigma}}^{(t)} \\ \boldsymbol{y}_{\hat{\sigma}}^{(t)} \end{pmatrix} \right\|_{t,\infty}. \tag{7}$$

For our choice of the learning rate $\eta \leq \frac{1}{50\|\mathcal{Q}_i\|_1}$, Lemma B.3 implies that $\mu^{(t)} \leq \frac{1}{2}$. By definition, we have

$$\boldsymbol{q}_{\hat{\sigma}}^{(t+1)} := \frac{\boldsymbol{y}_{\hat{\sigma}}^{(t+1)}}{\lambda_{\hat{\sigma}}^{(t+1)}} \leq \frac{(1+\mu^{(t)})\boldsymbol{y}_{\hat{\sigma}}^{(t)}}{(1-\mu^{(t)})\lambda_{\hat{\sigma}}^{(t)}} = \left(1 + \frac{2\mu^{(t)}}{1-\mu^{(t)}} \right) \boldsymbol{q}_{\hat{\sigma}}^{(t)} \leq (1+4\mu^{(t)})\boldsymbol{q}_{\hat{\sigma}}^{(t)},$$

where the last bound follows since $\mu^{(t)} \leq \frac{1}{2}$. That is,

$$\frac{\boldsymbol{q}_{\hat{\sigma}}^{(t+1)}}{\boldsymbol{q}_{\hat{\sigma}}^{(t)}} \leq (1+4\mu^{(t)}). \tag{8}$$

Similarly,

$$\boldsymbol{q}_{\hat{\sigma}}^{(t+1)} = \frac{\boldsymbol{y}_{\hat{\sigma}}^{(t+1)}}{\lambda_{\hat{\sigma}}^{(t+1)}} \geq \frac{1-\mu^{(t)}}{1+\mu^{(t)}} \frac{\boldsymbol{y}_{\hat{\sigma}}^{(t)}}{\lambda_{\hat{\sigma}}^{(t)}} = \left(1 - \frac{2\mu^{(t)}}{1+\mu^{(t)}}\right) \boldsymbol{q}_{\hat{\sigma}}^{(t)} \geq (1-2\mu^{(t)})\boldsymbol{q}_{\hat{\sigma}}^{(t)}.$$

Thus,

$$\frac{\boldsymbol{q}_{\hat{\sigma}}^{(t+1)}}{\boldsymbol{q}_{\hat{\sigma}}^{(t)}} \geq 1 - 2\mu^{(t)}. \tag{9}$$

As a result, the claim follows from (8) and (9). □

We are now ready to establish Proposition 3.1, restated below.

Proposition 3.1. *Fix any $\hat{\sigma} \in \Sigma_i^*$, and let $\mathrm{Reg}_{\hat{\sigma}}^T$ be the regret of $\mathfrak{R}_{\hat{\sigma}}$ up to time $T \geq 2$. For any $\eta \leq \frac{1}{256\|\mathcal{Q}_i\|_1}$, $\max\{0, \mathrm{Reg}_{\hat{\sigma}}^T\}$ can be upper bounded by*

$$\frac{2|\Sigma_i| \log T}{\eta} + 16\eta\|\mathcal{Q}_i\|_1^2 \sum_{t=1}^{T-1} \|\boldsymbol{U}_i^{(t+1)} - \boldsymbol{U}_i^{(t)}\|_\infty^2 - \frac{1}{512\eta} \sum_{t=1}^{T-1} \|\boldsymbol{q}_{\hat{\sigma}}^{(t+1)} - \boldsymbol{q}_{\hat{\sigma}}^{(t)}\|_{\boldsymbol{q}_{\hat{\sigma}}^{(t)},\infty}^2. \tag{3}$$

Proof. The claim follows directly from Proposition B.1 and Lemma B.2. □

Under the premise that $\mathfrak{R}_\triangle$ is also updated via `LRL-OFTRL`, similar reasoning yields the proof of Proposition 3.2.

Proposition 3.2. *Let $\mathrm{Reg}_\triangle^T$ be the regret of $\mathfrak{R}_\triangle$ up to time $T \geq 2$. For any $\eta_\triangle \leq \frac{1}{512|\Sigma_i|}$, $\max\{0, \mathrm{Reg}_\triangle^T\}$ can be upper bounded by*

$$\frac{2|\Sigma_i| \log T}{\eta_\triangle} + 16\eta_\triangle|\Sigma_i|^2 \sum_{t=1}^{T-1} \|\boldsymbol{u}_\triangle^{(t+1)} - \boldsymbol{u}_\triangle^{(t)}\|_\infty^2 - \frac{1}{512\eta_\triangle} \sum_{t=1}^{T-1} \|\boldsymbol{\lambda}_i^{(t+1)} - \boldsymbol{\lambda}_i^{(t)}\|_{\boldsymbol{\lambda}_i^{(t)},\infty}^2.$$

Proof. The argument is analogous to the proof of Proposition 3.1, leveraging the fact that $\|\boldsymbol{u}_\triangle^{(t)}\|_\infty = |\langle \boldsymbol{X}_{\hat{\sigma}}^{(t)}, \boldsymbol{U}_i^{(t)} \rangle| \leq \|\boldsymbol{X}_{\hat{\sigma}}^{(t)}\|_1 \|\boldsymbol{U}_i^{(t)}\|_\infty \leq 2|\Sigma_i|$, for any $\hat{\sigma} \in \Sigma_i^*$, by Cauchy-Schwarz inequality. □

Lemma B.4. *For any $t \in [\![T-1]\!]$ and $\eta_\triangle \leq \frac{1}{512|\Sigma_i|}$,*

$$\max_{\hat{\sigma} \in \Sigma_i^*} \left| 1 - \frac{\boldsymbol{\lambda}_i^{(t+1)}[\hat{\sigma}]}{\boldsymbol{\lambda}_i^{(t)}[\hat{\sigma}]} \right| \leq 200\eta_\triangle|\Sigma_i|.$$

Proof. The argument is analogous to Lemma B.2. □

Next, we combine all those local regret minimizers, namely $\mathfrak{R}_\triangle, (\mathfrak{R}_{\hat{\sigma}})_{\hat{\sigma} \in \Sigma_i^*}$, into a global regret minimizer $\mathfrak{R}_{\Psi_i}$ for the set Ψ_i via the regret circuit for the convex hull. Finally, we denote by $\mathfrak{R}$ the Ψ_i-regret minimizer derived from Algorithm 1, based on $\mathfrak{R}_{\Psi_i}$.

Proposition 3.3. *For any $T \in \mathbb{N}$,*

$$\max\{0, \mathrm{Reg}_{\Psi_i}^T\} \leq \max\{0, \mathrm{Reg}_\triangle^T\} + \sum_{\hat{\sigma} \in \Sigma_i^*} \max\{0, \mathrm{Reg}_{\hat{\sigma}}^T\}.$$

Proof. Using guarantee of the regret circuit for the convex hull [Farina et al., 2019a], we have

$$\mathrm{Reg}^T \leq \mathrm{Reg}^T_\triangle + \max_{\hat{\sigma} \in \Sigma_i^*} \mathrm{Reg}^T_{\hat{\sigma}},$$

where Reg^T is the external regret cumulated by $\mathfrak{R}_{\Psi_i}$ up to time T. But, by Theorem 2.1, this is equal to the Ψ_i-regret of $\mathfrak{R}$, constructed according to Algorithm 1. As a result,

$$\mathrm{Reg}^T_{\Psi_i} \leq \mathrm{Reg}^T_\triangle + \max_{\hat{\sigma} \in \Sigma_i^*} \mathrm{Reg}^T_{\hat{\sigma}},$$

In turn, this implies that

$$\begin{aligned}
\max\{0, \mathrm{Reg}^T_{\Psi_i}\} &\leq \max\left\{0, \mathrm{Reg}^T_\triangle + \max_{\hat{\sigma} \in \Sigma_i^*} \mathrm{Reg}^T_{\hat{\sigma}}\right\} \\
&\leq \max\{0, \mathrm{Reg}^T_\triangle\} + \max_{\hat{\sigma} \in \Sigma_i^*} \max\{0, \mathrm{Reg}^T_{\hat{\sigma}}\} \\
&\leq \max\{0, \mathrm{Reg}^T_\triangle\} + \sum_{\hat{\sigma} \in \Sigma_i^*} \max\{0, \mathrm{Reg}^T_{\hat{\sigma}}\},
\end{aligned}$$

where the last inequality follows from the fact that $\max\{0, \mathrm{Reg}^T_{\hat{\sigma}}\} \geq 0$, for any $\hat{\sigma} \in \Sigma_i^*$. □

B.2 Characterizing the Fixed Points

We recall that $(\boldsymbol{x}_i^{(t)})_{1 \leq t \leq T}$ denotes the sequence of fixed points produced by Algorithm 1—that is, the strategies produced by $\mathfrak{R}$. The next key result relates the deviation of the fixed points—in ℓ_1 norm—in terms of the *multiplicative deviation* of the transformations, assuming a particular rational function characterization of the fixed points.

Lemma 3.5. *Let $\boldsymbol{X}_i^{(t)} \in \mathbb{R}^D_{>0}$ be defined for every time $t \in \mathbb{N}$, for some $D \in \mathbb{N}$. Further, suppose that for every time $t \in \mathbb{N}$ and $\sigma \in \Sigma_i$,*

$$\boldsymbol{x}_i^{(t)}[\sigma] = \sum_{k=1}^{m} \frac{p_{\sigma,k}(\boldsymbol{X}_i^{(t)})}{q_{\sigma,k}(\boldsymbol{X}_i^{(t)})}, \tag{4}$$

for some multivariate polynomials $\{p_{\sigma,k}\}, \{q_{\sigma,k}\}$ with positive coefficients and maximum degree $\deg_i \in \mathbb{N}$. If

$$\max_{e \in [\![D]\!]} \left| 1 - \frac{\boldsymbol{X}_i^{(t+1)}[e]}{\boldsymbol{X}_i^{(t)}[e]} \right| \leq \frac{100}{256 \deg_i}, \tag{5}$$

it holds that

$$\|\boldsymbol{x}_i^{(t+1)} - \boldsymbol{x}_i^{(t)}\|_1 \leq (4 \|\mathcal{Q}_i\|_1 \deg_i) \max_{e \in [\![D]\!]} \left| 1 - \frac{\boldsymbol{X}_i^{(t+1)}[e]}{\boldsymbol{X}_i^{(t)}[e]} \right|.$$

Proof. Let us define

$$\mu^{(t)} := \max_{e \in [\![D]\!]} \left| 1 - \frac{\boldsymbol{X}_i^{(t+1)}[e]}{\boldsymbol{X}_i^{(t)}[e]} \right|. \tag{10}$$

By assumption, it holds that $\mu^{(t)} \leq \frac{100}{256 \deg_i} \leq \frac{1}{2 \deg_i}$. Further, suppose that

$$p_{\sigma,k} : \boldsymbol{X}_i \mapsto \sum_{\mathcal{T} \in \mathbb{T}_{\sigma,k}} C_{\mathcal{T}} \prod_{e \in \mathcal{T}} \boldsymbol{X}_i[e], \tag{11}$$

and

$$q_{\sigma,k} : \boldsymbol{X}_i \mapsto \sum_{\mathcal{T} \in \mathbb{T}'_{\sigma,k}} C_{\mathcal{T}} \prod_{e \in \mathcal{T}} \boldsymbol{X}_i[e], \tag{12}$$

for all $(\sigma, k) \in \Sigma_i \times [\![m]\!]$, where $C_{\mathcal{T}} > 0$ for any $\mathcal{T} \in \mathbb{T}_{\sigma,k}$ and $C_{\mathcal{T}} > 0$ for any $\mathcal{T} \in \mathbb{T}'_{\sigma,k}$. Here, $\mathcal{T}$ can be a multiset or an empty set. Then, for $(\sigma, k) \in \Sigma_i \times [\![m]\!]$,

$$\begin{aligned}
p_{\sigma,k}(\boldsymbol{X}_i^{(t+1)}) &= \sum_{\mathcal{T} \in \mathbb{T}_{\sigma,k}} C_{\mathcal{T}} \prod_{e \in \mathcal{T}} \boldsymbol{X}_i^{(t+1)}[e] \\
&\leq \sum_{\mathcal{T} \in \mathbb{T}_{\sigma,k}} C_{\mathcal{T}} \prod_{e \in \mathcal{T}} (1 + \mu^{(t)}) \boldsymbol{X}_i^{(t)}[e] \qquad (13) \\
&\leq (1 + \mu^{(t)})^{\deg_i} \sum_{\mathcal{T} \in \mathbb{T}_{\sigma,k}} C_{\mathcal{T}} \prod_{e \in \mathcal{T}} \boldsymbol{X}_i^{(t)}[e] \qquad (14) \\
&= (1 + \mu^{(t)})^{\deg_i} p_{\sigma,k}(\boldsymbol{X}_i^{(t)}) \\
&\leq (1 + 1.5\mu^{(t)} \deg_i) p_{\sigma,k}(\boldsymbol{X}_i^{(t)}), \qquad (15)
\end{aligned}$$

where (13) follows since $\boldsymbol{X}_i^{(t+1)}[e] \leq (1 + \mu^{(t)}) \boldsymbol{X}_i^{(t)}[e]$, for any $e \in [\![D]\!]$, by definition of $\mu^{(t)}$ in (10); (14) uses the fact that $|\mathcal{T}| \leq \deg$ for any $\mathcal{T} \in \mathbb{T}_{\sigma,k}$; and (15) follows since $(1 + \mu^{(t)})^{\deg_i} \leq \exp\{\mu^{(t)} \deg_i\} \leq 1 + 1.3\mu^{(t)} \deg_i$ for $\mu^{(t)} \leq \frac{1}{2 \deg_i}$. Similarly, for $(\sigma, k) \in \Sigma_i \times [\![m]\!]$, we get

$$\begin{aligned}
p_{\sigma,k}(\boldsymbol{X}_i^{(t+1)}) &= \sum_{\mathcal{T} \in \mathbb{T}_{\sigma,k}} C_{\mathcal{T}} \prod_{e \in \mathcal{T}} \boldsymbol{X}_i^{(t+1)}[e] \\
&\geq \sum_{\mathcal{T} \in \mathbb{T}_{\sigma,k}} C_{\mathcal{T}} \prod_{e \in \mathcal{T}} (1 - \mu^{(t)}) \boldsymbol{X}_i^{(t)}[e] \\
&\geq (1 - \mu^{(t)})^{\deg_i} p_{\sigma,k}(\boldsymbol{X}_i^{(t)}) \\
&\geq (1 - \mu^{(t)} \deg_i) p_{\sigma,k}(\boldsymbol{X}_i^{(t)}), \qquad (16)
\end{aligned}$$

where the last bound follows from Bernoulli's inequality. Analogous reasoning yields that for any $(\sigma, k) \in \Sigma_i \times [\![m]\!]$,

$$q_{\sigma,k}(\boldsymbol{X}_i^{(t+1)}) \leq (1 + 1.3\mu^{(t)} \deg_i) q_{\sigma,k}(\boldsymbol{X}_i^{(t)}), \tag{17}$$

and

$$q_{\sigma,k}(\boldsymbol{X}_i^{(t+1)}) \geq (1 - \mu^{(t)} \deg_i) q_{\sigma,k}(\boldsymbol{X}_i^{(t)}). \tag{18}$$

As a result, for $\sigma \in \Sigma_i$,

$$\begin{aligned}
\boldsymbol{x}_i^{(t+1)}[\sigma] - \boldsymbol{x}_i^{(t)}[\sigma] &= \sum_{k=1}^{m} \frac{p_{\sigma,k}(\boldsymbol{X}_i^{(t+1)})}{q_{\sigma,k}(\boldsymbol{X}_i^{(t+1)})} - \sum_{k=1}^{m} \frac{p_{\sigma,k}(\boldsymbol{X}_i^{(t)})}{q_{\sigma,k}(\boldsymbol{X}_i^{(t)})} \\
&\leq \sum_{k=1}^{m} \left(\frac{1 + 1.3\mu^{(t)} \deg_i}{1 - \mu^{(t)} \deg_i} \right) \frac{p_{\sigma,k}(\boldsymbol{X}^{(t)})}{q_{\sigma,k}(\boldsymbol{X}_i^{(t)})} - \sum_{k=1}^{m} \frac{p_{\sigma,k}(\boldsymbol{X}_i^{(t)})}{q_{\sigma,k}(\boldsymbol{X}_i^{(t)})} \qquad (19) \\
&\leq \left(1 + \frac{2.3\mu^{(t)} \deg_i}{1 - \mu^{(t)} \deg} \right) \sum_{k=1}^{m} \frac{p_{\sigma,k}(\boldsymbol{X}_i^{(t)})}{q_{\sigma,k}(\boldsymbol{X}_i^{(t)})} - \sum_{k=1}^{m} \frac{p_{\sigma,k}(\boldsymbol{X}_i^{(t)})}{q_{\sigma,k}(\boldsymbol{X}_i^{(t)})} \\
&= \frac{2.3\mu^{(t)} \deg_i}{1 - \mu^{(t)} \deg_i} \boldsymbol{x}_i^{(t)}[\sigma] \leq 4\mu^{(t)} \deg_i \boldsymbol{x}_i^{(t)}[\sigma]. \qquad (20)
\end{aligned}$$

where (19) uses (15) and (18), and (20) follows from the fact that $\mu^{(t)} \leq \frac{100}{256 \deg_i}$. Similarly, by (16) and (17),

$$\begin{aligned}
\boldsymbol{x}_i^{(t)}[\sigma] - \boldsymbol{x}_i^{(t+1)}[\sigma] &= \sum_{k=1}^{m} \frac{p_{\sigma,k}(\boldsymbol{X}_i^{(t)})}{q_{\sigma,k}(\boldsymbol{X}_i^{(t)})} - \sum_{k=1}^{m} \frac{p_{\sigma,k}(\boldsymbol{X}_i^{(t+1)})}{q_{\sigma,k}(\boldsymbol{X}_i^{(t+1)})} \\
&\leq 4\mu^{(t)} \deg_i \boldsymbol{x}_i^{(t)}[\sigma].
\end{aligned}$$

As a result, we conclude that

$$\|\boldsymbol{x}_i^{(t+1)} - \boldsymbol{x}_i^{(t)}\|_1 \leq 4\mu^{(t)} \deg_i \|\mathcal{Q}_i\|_1.$$

□

Lemma 3.5 makes the assumption that each polynomial in (4) has positive coefficients. While this might seem rather restrictive, we next show that there is a procedure that eliminates the negative monomials, as long as the involved variables are deriving from the sequence-form polytope. As a warm-up, we first establish this property for variables deriving from the simplex.

We note that the processes described in the proofs below are not meant to be algorithmic meaningful, but instead highlight the generality of Lemma 3.5. Indeed, the way one computes the fixed point should not be related to the rational function formula postulated in (4); for example, computing the stationary distribution of a Markov chain using the Markov chain tree theorem would make little sense, as it would require exponential time.

Proposition B.5. *Let $p : \boldsymbol{X} \mapsto \mathbb{R}$ be a non-constant multivariate polynomial of degree* deg $\in \mathbb{N}$ *such that $p(\mathbf{0}) = 0$. If $\boldsymbol{X} = (\boldsymbol{x}_1, \dots, \boldsymbol{x}_m)$ such that $\boldsymbol{x}_k \in \Delta^{d_k}$, for all $k \in [\![m]\!]$, p can be expressed as a combination of monomials with positive coefficients and degree at most* deg.

Proof. Let

$$p(\boldsymbol{X}) = \sum_{\mathcal{T} \in \mathbb{T}} C_{\mathcal{T}} \prod_{e \in \mathcal{T}} \boldsymbol{X}[e],$$

where $\mathbb{T}$ is a finite and nonempty set, and $\mathcal{T} \neq \emptyset$ and $C_{\mathcal{T}} \neq 0$ for all $\mathcal{T} \in \mathbb{T}$; the validity of such a formulation follows since, by assumption, $p(\mathbf{0}) = 0$ and p is non-constant. To establish the claim, we consider the following iterative algorithm.

First, if it happens that $C_{\mathcal{T}} > 0$, for all $\mathcal{T} \in \mathbb{T}$, the algorithm terminates. Otherwise, we take any monomial of the form $C_{\mathcal{T}} \prod_{e \in \mathcal{T}} \boldsymbol{X}[e]$ with $C_{\mathcal{T}} < 0$. Since $\mathcal{T} \neq \emptyset$, we might take $e \in \mathcal{T}$. Further, we let $\boldsymbol{X}[e] = \boldsymbol{x}_k[r]$, for some $k \in [\![m]\!], r \in [\![d_k]\!]$, where $\boldsymbol{x}_k \in \Delta^{d_k}$. As such, we have that $\boldsymbol{x}_k[r] = 1 - \sum_{r' \neq r} \boldsymbol{x}_k[r']$. Thus,

$$\begin{aligned} C_{\mathcal{T}} \prod_{e' \in \mathcal{T}} \boldsymbol{X}[e'] = C_{\mathcal{T}} &\prod_{e' \in \mathcal{T} \setminus \{e\}} \boldsymbol{X}[e'] \\ + \sum_{r' \neq r} (-C_{\mathcal{T}}) \boldsymbol{x}_k[r'] &\prod_{e' \in \mathcal{T} \setminus \{e\}} \boldsymbol{X}[e']. \end{aligned}$$

Here, by convention the product over an empty set is assumed to be 1. This step clearly cannot increase the degree of the polynomial. Now to analyze this iterative process, we consider as the potential function the sum of the degrees of all the negative monomials—monomials for which $C_{\mathcal{T}} < 0$. It should be evident that every step of the previous algorithm will decrease the potential function by one. Further, the previous step can always be applied as long as the potential function is not zero. As a result, given that $\mathbb{T}$ is finite, we conclude that after a finite number of iterations the potential function will be zero. Then, we will have that

$$p(\boldsymbol{X}) = \sum_{\mathcal{T} \in \mathbb{T}'} C_{\mathcal{T}} \prod_{e \in \mathcal{T}} \boldsymbol{X}[e] + C,$$

where $\mathcal{T} \neq \emptyset$ and $C_{\mathcal{T}} > 0$. But, given that $p(\mathbf{0}) = 0$, we conclude that $C = 0$, and the claim follows. □

Proposition B.6. *Let $p : \boldsymbol{X} \mapsto \mathbb{R}$ be a non-constant multivariate polynomial of degree* deg $\in \mathbb{N}$ *such that $p(\mathbf{0}) = 0$. If $\boldsymbol{X} = (\boldsymbol{q}_1, \dots, \boldsymbol{q}_m)$ such that $\boldsymbol{q}_k \in \mathcal{Q}^{d_k}$, for all $k \in [\![m]\!]$, p can be expressed as a combination of monomials with positive coefficients and degree at most* deg*.*

Proof. As in Proposition B.5, let

$$p(\boldsymbol{X}) = \sum_{\mathcal{T} \in \mathbb{T}} C_{\mathcal{T}} \prod_{e \in \mathcal{T}} \boldsymbol{X}[e],$$

where $\mathbb{T}$ is a finite and nonempty set, and $\mathcal{T} \neq \emptyset$ and $C_{\mathcal{T}} \neq 0$ for all $\mathcal{T} \in \mathbb{T}$. We consider the following algorithm.

First, if $C_{\mathcal{T}} > 0$, for all $\mathcal{T} \in \mathbb{T}$, the algorithm may terminate. In the contrary case, we consider any monomial $C_{\mathcal{T}} \prod_{e \in \mathcal{T}} \boldsymbol{X}[e]$ for which $C_{\mathcal{T}} < 0$. Further, take any $e \in \mathcal{T}$, which is possible since $\mathcal{T} \neq \emptyset$. Now let us assume that $\boldsymbol{X}[e] = \boldsymbol{q}_k[\sigma]$, for some $k \in [\![m]\!], \sigma = (j, a)$. By the sequence-form polytope constraints, we have

$$\boldsymbol{q}_k[\sigma] = \boldsymbol{q}_k[\sigma_j] - \sum_{a' \in \mathcal{A}_j \setminus \{a\}} \boldsymbol{q}_k[(j, a')].$$

Thus,

$$\begin{aligned} C_{\mathcal{T}} \prod_{e' \in \mathcal{T}} \boldsymbol{X}[e'] = C_{\mathcal{T}} \boldsymbol{q}_k[\sigma_j] &\prod_{e' \in \mathcal{T} \setminus \{e\}} \boldsymbol{X}[e'] \\ + \sum_{a' \neq a} (-C_{\mathcal{T}}) \boldsymbol{q}_k[(j, a')] &\prod_{e' \in \mathcal{T} \setminus \{e\}} \boldsymbol{X}[e']. \end{aligned}$$

This step clearly does not increase the degree of the polynomial. To construct a potential function, we will say that the *depth* of a monomial $\prod_{e \in \mathcal{T}} \boldsymbol{X}[e]$, for $\mathcal{T} \neq \emptyset$, is the sum of the depths of each $\boldsymbol{X}[e]$; more precisely, the depth of $\boldsymbol{q}_k[\sigma]$ is 0 if $\sigma = \varnothing$, or 1 plus the depth of $\boldsymbol{q}_k[\sigma_j]$ otherwise. Now we claim that the sum of the depths of the negative monomials is a proper potential function. Indeed, by construction every step reduces the potential by 1, while the previous step can always be applied when the potential function is not zero. As a result, given that $\mathbb{T}$ is finite, we conclude that after a finite number of iterations the potential function will be zero, which in turn implies that

$$p(\boldsymbol{X}) = \sum_{\mathcal{T} \in \mathbb{T}'} C_{\mathcal{T}} \prod_{e \in \mathcal{T}} \boldsymbol{X}[e] + C,$$

where $\mathcal{T} \neq \emptyset$ and $C_{\mathcal{T}} > 0$. But, since $p(\boldsymbol{0}) = 0$, it follows that $C = 0$, concluding the proof. □

Now we show that the fixed points associated with EFCCE and EFCE can be analyzed through the lens of Lemma 3.5, establishing Propositions 3.6 and 3.7.

Proposition 3.6. *Let $\phi_i^{(t)} \in \tilde{\Psi}_i$ be a transformation defined by $\boldsymbol{X}_i^{(t)} = (\boldsymbol{\lambda}_i^{(t)}, (\boldsymbol{q}_j^{(t)})_{j \in \mathcal{J}_i}) \in \mathbb{R}_{>0}^D$, for some $D \in \mathbb{N}$ and time $t \in \mathbb{N}$. The unique fixed point $\boldsymbol{x}_i^{(t)}$ of $\phi_i^{(t)}$ satisfies (4) with deg$_i \leq 2\mathfrak{D}_i$.*

Proof. Consider any coarse trigger deviation function $\phi_i^{(t)} = \sum_{j \in \mathcal{J}_i} \boldsymbol{\lambda}_i^{(t)}[j] \phi_{j \to \boldsymbol{q}_j^{(t)}}$, where $\boldsymbol{q}_j^{(t)} \in \mathcal{Q}_j$ (Claim A.4). Given that we are updating $\mathfrak{R}_\triangle$ using `LRL-OFTRL`, it follows that $\boldsymbol{\lambda}_i^{(t)}[j] > 0$ for any $j \in \mathcal{J}_i$. As a result, by [Anagnostides et al., 2022b, Theorem 5.1], the (unique) fixed point $\boldsymbol{x}_i^{(t)} \in \mathcal{Q}_i$ can be computed in a top-down fashion as follows.

$$\boldsymbol{x}_i^{(t)}[\sigma] = \frac{\sum_{j' \preceq j} \boldsymbol{\lambda}_i^{(t)}[j'] \boldsymbol{q}_{j'}^{(t)}[\sigma] \boldsymbol{x}_i^{(t)}[\sigma_{j'}]}{\sum_{j' \preceq j} \boldsymbol{\lambda}_i^{(t)}[j']}, \tag{21}$$

for any sequence $\sigma = (j, a) \in \Sigma_i^*$. We will prove the claim by induction. For the basis of the induction, we note that the empty sequence is trivially given by a 0-degree rational function with positive coefficients; namely, $\boldsymbol{x}_i[\varnothing] = \frac{1}{1}$.

Now for the inductive step, let us take any sequence $\sigma = (j, a) \in \Sigma_i^*$. We suppose that for any sequence $\sigma_{j'}$, for $j' \preceq j$, it holds that

$$\boldsymbol{x}_i^{(t)}[\sigma_{j'}] = \sum_{k=1}^{m_{\sigma_{j'}}} \frac{p_{\sigma_{j'},k}(\boldsymbol{X}_i^{(t)})}{q_{\sigma_{j'},k}(\boldsymbol{X}_i^{(t)})},$$

where $\{p_{\sigma_{j'},k}\}, \{q_{\sigma_{j'},k}\}$ are multivariate polynomials with positive coefficients and maximum degree at most $h \in \mathbb{N} \cup \{0\}$. Then, the term

$$\frac{\boldsymbol{\lambda}_i^{(t)}[j'] \boldsymbol{q}_{j'}^{(t)}[\sigma]}{\sum_{j' \preceq j} \boldsymbol{\lambda}_i^{(t)}[j']} \cdot \frac{p_{\sigma_{j'},k}(\boldsymbol{X}_i^{(t)})}{q_{\sigma_{j'},k}(\boldsymbol{X}_i^{(t)})}$$

is a rational function in $\boldsymbol{X}_i^{(t)}$ with positive coefficients and maximum degree at most $h + 2$. Hence, the term below is a sum of rational functions with positive coefficients and maximum degree at most $h + 2$:

$$\frac{\sum_{j' \preceq j} \boldsymbol{\lambda}_i^{(t)}[j'] \boldsymbol{q}_{j'}^{(t)}[\sigma] \boldsymbol{x}_i^{(t)}[\sigma_{j'}]}{\sum_{j' \preceq j} \boldsymbol{\lambda}_i^{(t)}[j']}$$

As a result, by (21) we conclude that

$$\boldsymbol{x}_i^{(t)}[\sigma] = \sum_{k=1}^{m_\sigma} \frac{p_{\sigma,k}(\boldsymbol{X}_i^{(t)})}{q_{\sigma,k}(\boldsymbol{X}_i^{(t)})},$$

where $\{p_{\sigma,k}\}, \{q_{\sigma,k}\}$ have positive coefficients and maximum degree $h+2$. This establishes the inductive step, concluding the proof. □

Next, for the proof of Proposition 3.7, we will need the following key refinement of the Markov chain tree theorem [Anagnostides et al., 2022b, Corollary A.8].

Theorem B.7 ([Anagnostides et al., 2022b])**.** *Let $\mathbf{M}$ be the transition matrix of a d-state Markov chain such that $\mathbf{M} = \boldsymbol{v}\mathbf{1}_d^\top + \mathbf{C}$, where $\mathbf{C} \in \mathbb{R}^{d\times d}_{>0}$ and $\boldsymbol{v} \in \mathbb{R}^d_{>0}$ has entries summing to $\lambda > 0$. Further, let $\boldsymbol{v} = \boldsymbol{r}/l$, for some $l > 0$. If $\boldsymbol{x} \in \Delta^d$ is the (unique) stationry distribution of $\mathbf{M}$, then for each $r \in [\![d]\!]$ there exist a nonempty and finite set F_r, and $F = \cup_{r=1}^d F_r$, and parameters $b_k \in \{0,1\}$, $0 \le p_k \le d-2$, $|S_k| = d - p_k - b_k - 1$, for each $k \in F_r$, such that the r-th coordinate of $\boldsymbol{w} \coloneqq l\boldsymbol{x}$ can be expressed as*

$$\boldsymbol{w}[r] = \frac{\sum_{k\in F_r} \lambda^{p_k+1}(\boldsymbol{r}[q_k])^{b_k} l^{1-b_k} \prod_{(a,b)\in S_k} \mathbf{C}[a,b]}{\sum_{k\in F} \lambda^{p_k+b_k} C_k \prod_{(a,b)\in S_k} \mathbf{C}[a,b]},$$

for each $r \in [\![d]\!]$, where $C_k = C_k(d) > 0$.

Let us also introduce the following terminology, borrowed from [Farina et al., 2021].

Definition B.8 ([Farina et al., 2021])**.** Let $J \subseteq \mathcal{J}_i$ be a subset of i's information sets. We say that J is a trunk of $\mathcal{J}_i$ if for all $j \in J$, all predecessors of j are also in J.

Definition B.9 ([Farina et al., 2021])**.** Let $\phi_i \in \Psi_i$ and J be a trunk of $\mathcal{J}_i$. We say that a vector $\boldsymbol{x}_i \in \mathbb{R}^{|\Sigma_i|}_{\ge 0}$ is a J-partial fixed point if it satisfies all the sequence-form constraints at all information sets $j \in \mathcal{J}$, and

$$\phi_i(\boldsymbol{x}_i)[\varnothing] = \boldsymbol{x}_i[\varnothing] = 1,$$
$$\phi_i(\boldsymbol{x}_i)[(j,a)] = \boldsymbol{x}_i[(j,a)], \quad \forall j \in \mathcal{J}, a \in \mathcal{A}_j.$$

Proposition 3.7. *Let $\phi_i^{(t)} \in \Psi_i$ be a transformation defined by $\boldsymbol{X}_i^{(t)} = (\boldsymbol{\lambda}_i^{(t)}, (\boldsymbol{q}_{\hat{\sigma}}^{(t)})_{\hat{\sigma}\in\Sigma_i^*}) \in \mathbb{R}^D_{>0}$, for some $D \in \mathbb{N}$ and time $t \in \mathbb{N}$. The (unique) fixed point $\boldsymbol{x}_i^{(t)}$ of $\phi_i^{(t)}$ satisfies (4) with $\deg_i \le 2\mathfrak{D}_i|\mathcal{A}_i|$, where $|\mathcal{A}_i| \coloneqq \max_{j\in\mathcal{J}_i} |\mathcal{A}_j|$.*

Proof. For the base of the induction, the claim trivially holds for $\boldsymbol{x}_i^{(t)}[\emptyset] = \frac{1}{1}$. For the inductive step, let us first define a vector $\boldsymbol{r}^{(t)} \in \mathbb{R}^{|\mathcal{A}_{j^*}|}_{\ge 0}$, so that $\boldsymbol{r}^{(t)}[a]$ is equal to

$$\sum_{j' \preceq \sigma_{j^*}} \sum_{a' \in \mathcal{A}_{j'}} \boldsymbol{\lambda}_i^{(t)}[(j',a')]\boldsymbol{q}^{(t)}_{(j',a')}[(j^*,a)]\boldsymbol{x}_i^{(t)}[(j',a')].$$

Further, we let $\mathbf{W}^{(t)} \in \mathbb{S}^{|\mathcal{A}_{j^*}|}$ be a stochastic matrix, so that for any $a_r, a_c \in \mathcal{A}_{j^*}$, $\mathbf{W}^{(t)}[a_r, a_c]$ is equal to

$$\frac{1}{\boldsymbol{x}_i^{(t)}[\sigma_{j^*}]}\boldsymbol{r}^{(t)}[a_r] + \boldsymbol{\lambda}_i^{(t)}[(j^*,a_c)]\boldsymbol{q}^{(t)}_{(j^*,a_c)}[(j^*,a_r)] + \left(1 - \sum_{\hat{\sigma} \preceq (j^*,a_c)} \boldsymbol{\lambda}_i^{(t)}[\hat{\sigma}]\right)\mathbf{1}\{a_r = a_c\},$$

By [Farina et al., 2021, Proposition 4.14], if $\boldsymbol{b}^{(t)} \in \Delta(\mathcal{A}_{j^*})$ is the (unique) stationary distribution of $\mathbf{W}^{(t)}$, extending by $\boldsymbol{x}_i^{(t)}[\sigma_{j^*}]\boldsymbol{b}^{(t)}$ at information set j^* yields a $(J \cup \{j^*\})$-partial fixed point (Definition B.9). To bound the increase in the degree of the rational function, we will use Theorem B.7. In particular, we define a matrix $\mathbf{C}^{(t)} \in \mathbb{R}^{|\mathcal{A}_{j^*}| \times |\mathcal{A}_{j^*}|}$, so that for any $a_r, a_c \in \mathcal{A}_{j^*}$,

$$\mathbf{C}^{(t)}[a_r, a_c] \coloneqq \boldsymbol{\lambda}_i^{(t)}[(j^*, a_c)]\boldsymbol{q}_{(j^*,a_c)}^{(t)}[(j^*, a_r)] + \left(1 - \sum_{\hat{\sigma} \preceq (j^*,a_c)} \boldsymbol{\lambda}_i^{(t)}[\hat{\sigma}]\right) \mathbf{1}\{a_r = a_c\}. \tag{22}$$

For a fixed $a_c \in \mathcal{A}_{j^*}$, we have

$$\sum_{a_r \in \mathcal{A}_{j^*}} \mathbf{C}^{(t)}[a_r, a_c] = \boldsymbol{\lambda}_i^{(t)}[(j^*, a_c)] + \sum_{\hat{\sigma} \not\preceq (j^*,a_c)} \boldsymbol{\lambda}_i^{(t)}[\hat{\sigma}], \tag{23}$$

where we used the fact that for any $a_c \in \mathcal{A}_{j^*}$, $\sum_{a_r \in \mathcal{A}_{j^*}} \boldsymbol{q}_{(j^*,a_c)}[(j^*, a_r)] = 1$ since $\boldsymbol{q}_{(j^*,a_c)} \in \mathcal{Q}_{j^*}$. Thus, from (23) we obtain that

$$1 - \sum_{a_r \in \mathcal{A}_{j^*}} \mathbf{C}^{(t)}[a_r, a_c] = \sum_{\hat{\sigma} \prec (j^*,a_c)} \boldsymbol{\lambda}_i^{(t)}[\hat{\sigma}]. \tag{24}$$

Now for the inductive step, suppose that for any information set $j' \preceq \sigma_{j^*}$ and $a' \in \mathcal{A}_{j'}$, the partial fixed point $\boldsymbol{x}_i^{(t)}[\sigma']$, with $\sigma' = (j', a')$, can be expressed as

$$\boldsymbol{x}_i^{(t)}[\sigma'] = \sum_{k=1}^{m_{\sigma'}} \frac{p_{\sigma',k}(\boldsymbol{X}_i^{(t)})}{q_{\sigma',k}(\boldsymbol{X}_i^{(t)})}, \tag{25}$$

where $\{p_{\sigma',k}\}, \{q_{\sigma',k}\}$ are multivariate polynomials with positive coefficients and maximum degree h. By (22), (24), the inductive hypothesis (25), and Theorem B.7, we conclude that for any $a \in \mathcal{A}_{j^*}$,

$$\boldsymbol{x}_i^{(t)}[(j^*, a)] = \sum_{k=1}^{m} \frac{p_{a,k}(\boldsymbol{X}_i^{(t)})}{q_{a,k}(\boldsymbol{X}_i^{(t)})},$$

where $\{p_{a,k}\}, \{q_{a,k}\}$ are multivariate polynomials with positive coefficients and maximum degree $h + 2|\mathcal{A}_{j^*}| \leq h + 2|\mathcal{A}_i|$. This concludes the inductive step, and the proof. □

Before we proceed, we note that while Propositions 3.1 and 3.2 and Lemmas B.2 and B.4 were stated for the construction relating to trigger deviations, those results readily apply for the construction relating to *coarse* trigger deviations as well; we omit the formal statements as they are almost identical to Propositions 3.1 and 3.2 and Lemmas B.2 and B.4.

In this context, combining Propositions 3.6 and 3.7 with Lemma 3.5 we arrive at the following conclusions.

Lemma B.10. *For any parameters* $\eta \leq \frac{1}{512\|\mathcal{Q}_i\|_1\mathfrak{D}_i}$, $\eta_\triangle \leq \frac{1}{1024|\Sigma_i|\mathfrak{D}_i}$, *and time* $t \in [\![T-1]\!]$,

$$\|\boldsymbol{x}_i^{(t+1)} - \boldsymbol{x}_i^{(t)}\|_1 \leq 8\|\mathcal{Q}_i\|_1\mathfrak{D}_i M(\boldsymbol{X}_i^{(t)}),$$

where $M(\boldsymbol{X}_i^{(t)})$ *is defined as*

$$\max\left\{\max_{j \in \mathcal{J}_i}\left|1 - \frac{\boldsymbol{\lambda}_i^{(t+1)}[j]}{\boldsymbol{\lambda}_i^{(t)}[j]}\right|, \max_{j \in \mathcal{J}_i}\max_{\sigma \in \Sigma_i}\left|1 - \frac{\boldsymbol{q}_j^{(t+1)}[\sigma]}{\boldsymbol{q}_j^{(t)}[\sigma]}\right|\right\}.$$

Proof. By Proposition 3.6, it follows that the fixed point $\boldsymbol{x}_i^{(t)}$ can be expressed, for any $\sigma \in \Sigma_i$, as

$$\boldsymbol{x}_i^{(t)}[\sigma] = \sum_{k=1}^{m} \frac{p_{\sigma,k}\left(\boldsymbol{\lambda}_i^{(t)}, (\boldsymbol{q}_j^{(t)})_{j \in \mathcal{J}_i}\right)}{q_{\sigma,k}\left(\boldsymbol{\lambda}_i^{(t)}, (\boldsymbol{q}_j^{(t)})_{j \in \mathcal{J}_i}\right)},$$

such that $\{p_{\sigma,k}\}, \{q_{\sigma,k}\}$ are multivariate polynomials in $\boldsymbol{X}_i^{(t)} = (\boldsymbol{\lambda}_i^{(t)}, (\boldsymbol{q}_j^{(t)})_{j \in \mathcal{J}_i})$ with positive coefficients and maximum degree $\deg_i \coloneqq 2\mathfrak{D}_i$. As a result, similarly to Lemmas B.2 and B.4, it follows that

$$\max_{\sigma \in \Sigma_j} \left| 1 - \frac{\boldsymbol{q}_j^{(t+1)}[\sigma]}{\boldsymbol{q}_j^{(t)}[\sigma]} \right| \leq 100\eta \|\mathcal{Q}_i\|_1 \leq \frac{100}{256 \deg_i},$$

for any $j \in \mathcal{J}_i$, and

$$\max_{j \in \mathcal{J}_i} \left| 1 - \frac{\boldsymbol{\lambda}_i^{(t+1)}[j]}{\boldsymbol{\lambda}_i^{(t)}[j]} \right| \leq 200\eta_\triangle |\Sigma_i| \leq \frac{100}{256 \deg_i}.$$

As a result, the claim follows from Lemma 3.5. □

Lemma 3.8. *Consider any parameters* $\eta \leq \frac{1}{256\|\mathcal{Q}_i\|_1 \deg_i}$ *and* $\eta_\triangle \leq \frac{1}{512|\Sigma_i| \deg_i}$, *where* $\deg_i \coloneqq 2|\mathcal{A}_i|\mathfrak{D}_i$. *Then, for any time* $t \in [\![T-1]\!]$,

$$\|\boldsymbol{x}_i^{(t+1)} - \boldsymbol{x}_i^{(t)}\|_1 \leq 8\|\mathcal{Q}_i\|_1 |\mathcal{A}_i| \mathfrak{D}_i M(\boldsymbol{X}_i^{(t)}),$$

where $M(\boldsymbol{X}_i^{(t)})$ *is defined as*

$$\max\left\{ \max_{\hat{\sigma} \in \Sigma_i^*} \left| 1 - \frac{\boldsymbol{\lambda}_i^{(t+1)}[\hat{\sigma}]}{\boldsymbol{\lambda}_i^{(t)}[\hat{\sigma}]} \right|, \max_{\hat{\sigma} \in \Sigma_i^*} \max_{\sigma \in \Sigma_i} \left| 1 - \frac{\boldsymbol{q}_{\hat{\sigma}}^{(t+1)}[\sigma]}{\boldsymbol{q}_{\hat{\sigma}}^{(t)}[\sigma]} \right| \right\}.$$

Proof. By Proposition 3.7, it follows that the fixed point $\boldsymbol{x}_i^{(t)}$ can be expressed, for any $\sigma \in \Sigma_i$, as

$$\boldsymbol{x}_i^{(t)}[\sigma] = \sum_{k=1}^{m} \frac{p_{\sigma,k}\left(\boldsymbol{\lambda}_i^{(t)}, (\boldsymbol{q}_{\hat{\sigma}}^{(t)})_{\hat{\sigma} \in \Sigma_i^*}\right)}{q_{\sigma,k}\left(\boldsymbol{\lambda}_i^{(t)}, (\boldsymbol{q}_{\hat{\sigma}}^{(t)})_{\hat{\sigma} \in \Sigma_i^*}\right)},$$

such that $\{p_{\sigma,k}\}, \{q_{\sigma,k}\}$ are multivariate polynomials in $\boldsymbol{X}_i^{(t)} = (\boldsymbol{\lambda}_i^{(t)}, (\boldsymbol{q}_{\hat{\sigma}}^{(t)})_{\hat{\sigma} \in \Sigma_i^*})$ with positive coefficients and maximum degree $\deg_i \coloneqq 2|\mathcal{A}_i|\mathfrak{D}_i$. As a result, in light of Lemmas B.2 and B.4,

$$\max_{\sigma \in \Sigma_j} \left| 1 - \frac{\boldsymbol{q}_{\hat{\sigma}}^{(t+1)}[\sigma]}{\boldsymbol{q}_{\hat{\sigma}}^{(t)}[\sigma]} \right| \leq 100\eta \|\mathcal{Q}_i\|_1 \leq \frac{100}{256 \deg_i},$$

for any $\hat{\sigma} = (j, a) \in \Sigma_i^*$, and

$$\max_{\hat{\sigma} \in \Sigma_i^*} \left| 1 - \frac{\boldsymbol{\lambda}_i^{(t+1)}[\hat{\sigma}]}{\boldsymbol{\lambda}_i^{(t)}[\hat{\sigma}]} \right| \leq 200\eta_\triangle |\Sigma_i| \leq \frac{100}{256 \deg_i}.$$

As a result, the claim follows from Lemma 3.5. □

B.3 Completing the Proof

Finally, here we combine all of the previous ingredients to complete the proof of Theorem 3.10.

Proposition B.11. *Let* $\eta \leq \frac{1}{256|\Sigma_i|^{1.5}}$ *and* $\eta_\triangle \leq \frac{1}{512|\Sigma_i|^{2.5}}$. *Then, for any* $T \geq 2$,

$$\begin{aligned}
\max\{0, \mathrm{Reg}^T_{\Psi_i}\} \leq{}& \frac{2|\Sigma_i|^2 \log T}{\eta} + \frac{2|\Sigma_i| \log T}{\eta_\triangle} \\
&+ (32\eta|\Sigma_i||\mathcal{Q}_i|^2 + 256\eta_\triangle|\Sigma_i|^4) \sum_{t=1}^{T-1} \|\boldsymbol{u}_i^{(t+1)} - \boldsymbol{u}_i^{(t)}\|_\infty^2 + (32\eta|\Sigma_i||\mathcal{Q}_i|^2 + 256\eta_\triangle|\Sigma_i|^4) \sum_{t=1}^{T-1} \|\boldsymbol{x}_i^{(t+1)} - \boldsymbol{x}_i^{(t)}\|_\infty^2 \\
&- \frac{1}{512\eta_\triangle} \sum_{t=1}^{T-1} \|\boldsymbol{\lambda}_i^{(t+1)} - \boldsymbol{\lambda}_i^{(t)}\|^2_{\boldsymbol{\lambda}_i^{(t)},\infty} - \frac{1}{1024\eta} \sum_{\hat{\sigma}\in\Sigma_i} \sum_{t=1}^{T-1} \|\boldsymbol{q}_{\hat{\sigma}}^{(t+1)} - \boldsymbol{q}_{\hat{\sigma}}^{(t)}\|_{\boldsymbol{q}_{\hat{\sigma}}^{(t)},\infty}.
\end{aligned}$$

Proof. Fix any $t \in [\![T-1]\!]$. By definition of $\boldsymbol{U}_i^{(t)}$ (Line 6),

$$\begin{aligned}
\|\boldsymbol{U}_i^{(t+1)} - \boldsymbol{U}_i^{(t)}\|_\infty^2 &\leq \|\boldsymbol{u}_i^{(t+1)} \otimes \boldsymbol{x}_i^{(t+1)} - \boldsymbol{u}_i^{(t)} \otimes \boldsymbol{x}_i^{(t)}\|_\infty^2 \\
&\leq 2\|\boldsymbol{u}_i^{(t+1)} \otimes (\boldsymbol{x}_i^{(t+1)} - \boldsymbol{x}_i^{(t)})\|_\infty^2 + 2\|\boldsymbol{x}_i^{(t)} \otimes (\boldsymbol{u}_i^{(t+1)} - \boldsymbol{u}_i^{(t)})\|_\infty^2 \qquad (26)\\
&= 2\|\boldsymbol{u}_i^{(t+1)}\|_\infty^2 \|\boldsymbol{x}_i^{(t+1)} - \boldsymbol{x}_i^{(t)}\|_\infty^2 + 2\|\boldsymbol{x}_i^{(t)}\|_\infty^2 \|\boldsymbol{u}_i^{(t+1)} - \boldsymbol{u}_i^{(t)}\|_\infty^2 \qquad (27)\\
&\leq 2\|\boldsymbol{x}_i^{(t+1)} - \boldsymbol{x}_i^{(t)}\|_\infty^2 + 2\|\boldsymbol{u}_i^{(t+1)} - \boldsymbol{u}_i^{(t)}\|_\infty^2, \qquad (28)
\end{aligned}$$

where (26) follows from the triangle inequality for the $\|\cdot\|_\infty$ norm, as well as Young's inequality; (27) uses the fact that $\|\boldsymbol{x} \otimes \boldsymbol{u}\|_\infty = \|\boldsymbol{x}\|_\infty \|\boldsymbol{u}\|_\infty$, for any vectors $\boldsymbol{x}, \boldsymbol{u}$; and (28) follows from the assumption that $\|\boldsymbol{u}_i^{(t)}\|_\infty, \|\boldsymbol{x}_i^{(t)}\|_\infty \leq 1$. Similarly, for any $t \in [\![T-1]\!]$,

$$\begin{aligned}
\|\boldsymbol{u}_\triangle^{(t+1)} - \boldsymbol{u}_\triangle^{(t)}\|_\infty^2 &= |\langle \boldsymbol{X}_{\hat{\sigma}}^{(t+1)}, \boldsymbol{U}_i^{(t+1)}\rangle - \langle \boldsymbol{X}_{\hat{\sigma}}^{(t)}, \boldsymbol{U}_i^{(t)}\rangle|^2 \qquad (29)\\
&\leq 2|\langle \boldsymbol{X}_{\hat{\sigma}}^{(t+1)}, \boldsymbol{U}_i^{(t+1)} - \boldsymbol{U}_i^{(t)}\rangle|^2 + 2|\langle \boldsymbol{U}_i^{(t)}, \boldsymbol{X}_{\hat{\sigma}}^{(t+1)} - \boldsymbol{X}_{\hat{\sigma}}^{(t)}\rangle|^2 \qquad (30)\\
&\leq 8|\Sigma_i|^2 \|\boldsymbol{U}_i^{(t+1)} - \boldsymbol{U}_i^{(t+1)}\|_\infty^2 + 2\|\boldsymbol{q}_{\hat{\sigma}}^{(t+1)} - \boldsymbol{q}_{\hat{\sigma}}^{(t)}\|_1^2, \qquad (31)
\end{aligned}$$

for some $\hat{\sigma} \in \Sigma_i^*$, where (29) follows from the definition of $\boldsymbol{u}_\triangle^{(t)}$; (30) uses Young's inequality; and (31) uses the Cauchy-Schwarz inequality, along with the fact that $\|\boldsymbol{U}_i^{(t)}\|_\infty \leq 1$ and $\|\boldsymbol{X}_{\hat{\sigma}}^{(t)}\|_1 \leq 2|\Sigma_i|$. Further, for any $\hat{\sigma} \in \Sigma_i^*$, $\eta \leq \frac{1}{256|\Sigma_i|^{1.5}}$ and $\eta_\triangle \leq \frac{1}{512|\Sigma_i|^{2.5}}$,

$$\begin{aligned}
&- \frac{1}{1024\eta} \|\boldsymbol{q}_{\hat{\sigma}}^{(t+1)} - \boldsymbol{q}_{\hat{\sigma}}^{(t)}\|^2_{\boldsymbol{q}_{\hat{\sigma}}^{(t)},\infty} + 32\eta_\triangle|\Sigma_i|^2 \|\boldsymbol{q}_{\hat{\sigma}}^{(t+1)} - \boldsymbol{q}_{\hat{\sigma}}^{(t)}\|_1^2 \\
&\leq \left(-\frac{1}{1024\eta} + 32\eta_\triangle|\Sigma_i|^4\right) \|\boldsymbol{q}_{\hat{\sigma}}^{(t+1)} - \boldsymbol{q}_{\hat{\sigma}}^{(t)}\|_\infty^2 \\
&\leq \left(-\frac{|\Sigma_i|^{1.5}}{4} + \frac{|\Sigma_i|^{1.5}}{16}\right) \|\boldsymbol{q}_{\hat{\sigma}}^{(t+1)} - \boldsymbol{q}_{\hat{\sigma}}^{(t)}\|_\infty^2 \leq 0. \qquad (32)
\end{aligned}$$

As a result, the proof follows from Propositions 3.1 to 3.3, (28), (31), and (32). □

As a result, we are now ready to establish Corollary 3.9, the statement of which is recalled below.

Corollary 3.9. *Suppose that* $\eta \leq \frac{1}{2^{12}|\Sigma_i|^{1.5}\|\mathcal{Q}_i\|_1 \deg_i}$ *and* $\eta_\triangle = \frac{1}{2|\Sigma_i|}\eta$*, where* $\deg_i := 2|\mathcal{A}_i|\mathfrak{D}_i$*. For any* $T \geq 2$*,* $\max\{0, \mathrm{Reg}_{\Psi_i}^T\}$ *can be upper bounded by*

$$\frac{8|\Sigma_i|^2 \log T}{\eta} + 256\eta|\Sigma_i|^3 \sum_{t=1}^{T-1} \|\boldsymbol{u}_i^{(t+1)} - \boldsymbol{u}_i^{(t)}\|_\infty^2 - \frac{1}{2^{15}\eta \deg_i^2 \|\mathcal{Q}_i\|_1^2} \sum_{t=1}^{T-1} \|\boldsymbol{x}_i^{(t+1)} - \boldsymbol{x}_i^{(t)}\|_1^2.$$

Proof. By Lemma 3.8,

$$\frac{1}{512\eta_\triangle} \sum_{t=1}^{T-1} \|\boldsymbol{\lambda}_i^{(t+1)} - \boldsymbol{\lambda}_i^{(t)}\|_{\boldsymbol{\lambda}_i^{(t)},\infty}^2 + \frac{1}{1024\eta} \sum_{\hat{\sigma}\in\Sigma_i} \sum_{t=1}^{T-1} \|\boldsymbol{q}_{\hat{\sigma}}^{(t+1)} - \boldsymbol{q}_{\hat{\sigma}}^{(t)}\|_{\boldsymbol{q}_{\hat{\sigma}}^{(t)},\infty} \\ \geq \frac{1}{2^{14}\eta\|\mathcal{Q}_i\|_1^2 \deg_i^2} \|\boldsymbol{x}_i^{(t+1)} - \boldsymbol{x}_i^{(t)}\|_1^2.$$

Thus, the proof follows directly from Proposition B.11 since for any $t \in [\![T-1]\!]$,

$$\left(256\eta|\Sigma_i|^3 - \frac{1}{2^{15}\eta\|\mathcal{Q}_i\|_1^2 \deg_i^2}\right) \|\boldsymbol{x}_i^{(t+1)} - \boldsymbol{x}_i^{(t)}\|_1^2 \leq 0,$$

for any $\eta \leq \frac{1}{2^{12}|\Sigma_i|^{1.5}\|\mathcal{Q}_i\|_1 \deg_i}$. □

So far we have performed the analysis from the perspective of a fixed player $i \in [\![n]\!]$, while being oblivious to the mechanism that produces the sequence of utilities $(\boldsymbol{u}_i^{(t)})_{1\leq t\leq T}$. Having established the RVU bound of Corollary 3.9, we are ready to show that when *all* players employ our learning dynamics, the second-order path lengths are bounded by $O(\log T)$. (In what follows, we tacitly assume that each player uses $\eta_\triangle := \frac{1}{2|\Sigma_i|}\eta$, in accordance to Corollary 3.9.)

Theorem B.12. *Suppose that each player* $i \in [\![n]\!]$ *uses learning rate* $\eta \leq \frac{1}{2^{12}(n-1)|\Sigma|^{1.5}\|\mathcal{Q}\|_1|\mathcal{Z}|\deg}$*, where* $\deg = 2|\mathcal{A}|\mathfrak{D}$*. Then, for any* $T \geq 2$*,*

$$\sum_{t=1}^{T-1} \sum_{i=1}^{n} \|\boldsymbol{x}_i^{(t+1)} - \boldsymbol{x}_i^{(t)}\|_1^2 \leq 2^{19} n|\Sigma|^2 \|\mathcal{Q}\|_1^2 \deg^2 \log T.$$

Proof. For any time $t \in [\![T-1]\!]$ and player $i \in [\![n]\!]$,

$$\|\boldsymbol{u}_i^{(t+1)} - \boldsymbol{u}_i^{(t)}\|_\infty^2 \leq (n-1)|\mathcal{Z}|^2 \sum_{i'\neq i} \|\boldsymbol{x}_{i'}^{(t+1)} - \boldsymbol{x}_{i'}^{(t)}\|_1^2,$$

by [Anagnostides et al., 2022b, Claim 4.16]. Further, for $\eta \leq \frac{1}{2^{12}(n-1)|\Sigma|^{1.5}\|\mathcal{Q}\|_1|\mathcal{Z}|\deg}$,

$$\left(256\eta(n-1)^2|\Sigma|^3|\mathcal{Z}|^2 - \frac{1}{2^{16}\eta\|\mathcal{Q}\|_1^2 \deg^2}\right) \leq 0.$$

As a result, using Corollary 3.9, $\sum_{i=1}^n \max\{0, \mathrm{Reg}_{\Psi_i}^T\}$ can be upper bounded by

$$\frac{8n|\Sigma|^2 \log T}{\eta} - \frac{1}{2^{16}\eta \deg^2 \|\mathcal{Q}\|_1^2} \sum_{i=1}^{n} \sum_{t=1}^{T-1} \|\boldsymbol{x}_i^{(t+1)} - \boldsymbol{x}_i^{(t)}\|_1^2.$$

But, given that $\sum_{i=1}^{n} \max\{0, \mathrm{Reg}_{\Psi_i}^T\} \geq 0$, we conclude that

$$\sum_{i=1}^{n} \sum_{t=1}^{T-1} \|\boldsymbol{x}_i^{(t+1)} - \boldsymbol{x}_i^{(t)}\|_1^2 \leq 2^{19} n |\Sigma|^2 \deg^2 \|\mathcal{Q}\|_1^2 \log T.$$

□

We now arrive at Theorem 3.10, which is restated below with the precise parameterization.

Corollary B.13. *Suppose that all players employ Algorithm 1 instantiated with* `LRL-OFTRL` *for all local regret minimizers,* $\mathfrak{R}_\triangle$ *and* $\{\mathfrak{R}_{\hat{\sigma}}\}_{\hat{\sigma}}$*, with* $\eta = \frac{1}{2^{13}(n-1)|\Sigma|^{1.5}\|\mathcal{Q}\|_1|\mathcal{Z}||\mathcal{A}|\mathfrak{D}}$ *and* $\eta_\triangle = \frac{1}{2|\Sigma_i|}\eta$*. Then, the trigger regret of each player* $i \in [\![n]\!]$ *after* T *repetitions will be bounded as*

$$\mathrm{Reg}_{\Psi_i}^T \leq 2^{17} n |\Sigma|^{3.5} \|\mathcal{Q}\|_1 |\mathcal{Z}||\mathcal{A}|\mathfrak{D} \log T.$$

Proof. This follows directly from Corollary 3.9 and Theorem B.12. □

Corollary B.14. *Suppose that all players employ Algorithm 1 instantiated with* `LRL-OFTRL` *for all local regret minimizers,* $\mathfrak{R}_\triangle$ *and* $\{\mathfrak{R}_j\}_j$*, with* $\eta = \frac{1}{2^{13}(n-1)|\Sigma|^{1.5}\|\mathcal{Q}\|_1|\mathcal{Z}|\mathfrak{D}}$ *and* $\eta_\triangle = \frac{1}{2|\Sigma_i|}\eta$*. Then, the trigger regret of each player* $i \in [\![n]\!]$ *after* T *repetitions will be bounded as*

$$\mathrm{Reg}_{\Psi_i}^T \leq 2^{17} n |\Sigma|^{3.5} \|\mathcal{Q}\|_1 |\mathcal{Z}|\mathfrak{D} \log T. \tag{33}$$

Proof. The proof is analogous to Corollary B.13. □

We remark that for *coarse* trigger regret, our bound (33) is loose, as the analysis is not optimized to handle coarse trigger deviation functions; instead, Corollary B.14 follows the construction of trigger deviations, with the exception of using Lemma B.10 in order to obtain a slightly improved RVU bound. Further refining Corollary B.14 was not within our scope.

C Description of the Game Instances

In this section, to keep our paper self-contained, we describe the games we used in our experiments (Section 4), as well as the precise parameterization for each instance.

Kuhn poker *Kuhn poker* is a simple poker variant studied by Kuhn [1953]. For simplicity, below we describe the 2-player version of Kuhn poker; the 3-player version we consider in our experiments is analogous.

In Kuhn poker each player initially submits an ante worth of 1 in the pot. Then, each player is privately dealt one card from a deck of r unique cards—or *ranks*; in our experiments we used $r = 3$. Next, a single round of betting occurs: First, player 1 gets to decide either check or bet. Then,

- If player 1 checked, the second player can either check or raise.
    - If player 2 also checked, a "showdown" occurs, meaning that the player with the highest card wins the pot, thereby terminating the game.
    - On the other hand, if player 2 raised, player 1 can either fold or call; in the former case player 2 wins the pot, while in the latter a showdown follows.

- If player 1 raised, player 2 can either fold or call.
  - If player 2 folded, then player 1 wins the pot, while
  - if player 2 called, a showdown occurs.

Sheriff *Sheriff* [Farina et al., 2019c] is a 2-player bargaining game inspired by the board game "Sheriff of Nottingham." Initially, player 1 (or the "Smuggler") secretly loads his cargo with $m \in \{0, 1, \dots, m_{\max}\}$ illegal items. The game then proceeds for r bargaining rounds. In each round,

- the Smuggler first gets to decide a bribe amount b in $\{0, 1, \dots, b_{\max}\}$. This amount also becomes available to player 2 (the "Sheriff"), although the smuggler does not transfer than amount unless it is the ultimate round.
- The Sheriff then decides whether to accept the bribe.
  - If the Sheriff accepts the bribe of value b, the smuggler gets a payoff of $p \cdot m - b$, while Sheriff receives a payoff of b.
  - In the contrary case, Sheriff decides whether to inspect the cargo.
    - If the Sheriff does not inspect the cargo, the Smuggler receives a payoff of $v \cdot m$, while the Sheriff gets 0 utility;
    - Otherwise, if the Sheriff detects illegal items, the Smuggler must pay the Sheriff an amount of $p \cdot m$, while if no illegal items were loaded, the Sheriff has to compensate the Smuggler with a utility of s.

In our experiments, we use the baseline version of Sheriff, wherein $v = 5, p = 1, s = 1, m_{\max} = 5, b_{\max} = 2$, and $r = 2$.

Goofspiel Goofspiel is a 2-player card game introduced by Ross [1971]. The game is based on three identical decks of r cards each, with values ranging from 1 to r; we use $r = 3$ in our experiments. Initially, each player is dealt a full deck, while the third deck (the "prize" deck) is faced down on the board after being shuffled. In each round, the topmost card from the prize deck is revealed. Then, each player privately selects a card from their hand with the goal of winning the card that was revealed from the prize deck. The players' selected cards are revealed simultaneously, and the card with the highest value prevails; in case of a tie, the prize card is discarded. This tie-breaking mechanism makes the game general-sum. Finally, the score of each player is the sum of the values of the prize cards that player has won.